Ontology of the Social

Agency, Attitudes and Ethics

Edited by Bhaskarjit Neog



ROUTLEDGE

Ontology of the Social

Agency, Attitudes and Ethics

Edited by Bhaskarjit Neog



ROUTLEDGE

OceanofPDF.com

“A stellar collection of some of the best work by some of the best contributors to debates in social metaphysics and collective ethics.”

Arto Laitinen, *Tampere University*

“This is an excellent volume of essays on social ontology. It delves into the metaphysics, moral psychology, and practical reasoning underlying collective and institutional agency, and explores implications for ethics and moral responsibility. Central figures and new voices guide readers into the diversity of topics and approaches falling under the heading of social ontology. But it doesn’t just serve as an introduction to these issues; there is much material here crucial for experts keeping abreast of developments in the field.”

Abraham Sesshu Roth, *Ohio State University*

OceanofPDF.com

Ontology of the Social

Ontology of the Social: Agency, Attitudes, and Ethics brings together leading and emerging voices in analytic social ontology to address some of the most pressing philosophical issues concerning social reality, and its nature and features. The chapters in this volume engage with a set of core concerns that have significantly shaped contemporary debates in social ontology and collective ethics. Some of the themes that these essays explore are: collectivity, social groups, plural and collective action, agency, we-attitudes, and collective responsibility. Moving beyond traditional individualist frameworks, the contributors investigate how recent developments in these themes have transformed our understanding of social reality and influenced broader philosophical discussions about human action, shared agency, and normativity.

Combining original scholarship with interdisciplinary perspectives, this volume will appeal to researchers and students working in social ontology, ethics, social and political philosophy, and philosophy of mind and action.

Bhaskarjit Neog teaches philosophy at Jawaharlal Nehru University, Delhi, India. His areas of interest are social ontology, ethics, and moral and political philosophy. His 2024 book, *What Responsibility? Whose Responsibility?*, explores the possibility of a non-individualist explanation of collective responsibility.

OceanofPDF.com

Ontology of the Social

Agency, Attitudes, and Ethics

Edited by Bhaskarjit Neog

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

OceanofPDF.com

Designed cover image: Bhaskarjit Neog

First published 2027
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2027 selection and editorial matter, Bhaskarjit Neog; individual chapters, the contributors

The right of Bhaskarjit Neog to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

For Product Safety Concerns and Information please contact our EU representative GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH, Kaufingerstraße 24, 80331 München, Germany.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-041-00764-7 (hbk)
ISBN: 978-1-041-01044-9 (pbk)
ISBN: 978-1-003-61282-7 (ebk)

DOI: 10.4324/9781003612827

Typeset in Sabon
by Deanta Global Publishing Services, Chennai, India

OceanofPDF.com

For those young lives tirelessly searching for light amid the ruins of conflict

OceanofPDF.com

Contents

List of Contributors
Acknowledgments

Introduction: Ontology of the Social: Agency, Attitudes, and Ethics

1 Realism, Constructionism, and Social Change
DAVID P. SCHWEIKARD

2 Institutions and Normativity: Needs, Joint Action, and Reciprocity
SEUMAS MILLER

3 Uptake: A Plural Action View
HANS BERNHARD SCHMID

4 Rethinking Individual and Group Responsibility in the Light of a Normative Account of Personal Identity
CAROL ROVANE

5 Collective Obligations Without the ‘Quantum of Blame’
BILL WRINGE

6 Moral Responsibility of Random Collectives

BHASKARJIT NEOG

7 Legitimate Firms
STEPHANIE COLLINS

8 The Problem of Overlapping Collectives
LUKE ROELOFS

9 Team Reasoning Without Assurance: Morality as a (Largely) Non-Coordinating Practice
JEREMY RANDEL KOONS

10 Social Reflective Equilibrium and Moral Progress
SUSHRUTH RAVISH

11 On the Content and Role of We-intentions
OLLE BLOMBERG

12 The Second-person Perspective on the First-person Plural
BIJOY H. BORUAH

13 First-Person Perspective and Collective Intentionality: A Relook into the Relationship between *I* and *We*
MANIDIPA SEN

Index

OceanofPDF.com

Contributors

Olle Blomberg is researcher in practical philosophy at the University of Gothenburg and HCAS-RJ fellow at the Helsinki Collegium of Advanced Studies 2025–2026. His research focuses on agency and moral responsibility, especially in social and collective contexts.

Bijoy H. Boruah is a Professor of philosophy at Indian Institute of Technology Jammu, India. He specializes in philosophy of mind, philosophy of aesthetics, philosophy of literature, metaphysics of the self. He has authored the book *Fiction and Emotion: A Study in the Philosophy of Mind* (Clarendon Press, Oxford, 1988). Some of his latest publications include "Reality and Intelligibility," in Partha Ghose (ed), *Tagore, Einstein and the Nature of Reality: Literary and Philosophical Dimensions* (Routledge, 2019), and "The Impersonal Subjectivity of Aesthetic Emotion," in Arindam Chakraborti (ed), *Indian Aesthetics and the Philosophy of Art* (Bloomsbury, 2016).

Stephanie Collins is Associate Professor of Philosophy at Monash University. She mostly works on collective agency and responsibility. Her books are *The Core of Care Ethics* (2015), *Group Duties: Their Existence and Their Implications for Individuals* (2019), and *Organizations as Wrongdoers: from Ontology to Morality* (2023).

Jeremy Koons, Professor of Philosophy at Georgetown University in Qatar, is the author of *The Ethics of Wilfrid Sellars* (Routledge, 2019) and the co-editor (with Kyle Ferguson) of *Wilfrid Sellars, The Metaphysics of Practice: Writings on Action, Community, and Obligation* (OUP, 2023). He has written, co-written, and co-edited three additional books, and has

published two dozen articles and book chapters on metaethics, epistemology, and philosophy of religion.

Seumas Miller is Professor of Philosophy at the Australian Graduate School of Policing and Security at Charles Sturt University and Distinguished Research Fellow at the Oxford Uehiro Institute at the University of Oxford. He is the author of over 200 academic articles and 22 books, including *Social Action: A Teleological Account* (Cambridge University Press, 2001), *The Moral Foundations of Social Institutions* (Cambridge University Press, 2010), and *Cybersecurity, Ethics and Collective Responsibility* (with T Bossomaier) (Oxford University Press, 2024).

Sushruth Ravish is an Assistant Professor of Philosophy in the Department of HSS at IIT Kanpur, India. He completed his PhD at IIT Bombay, where he was awarded the Naik and Rastogi Prize for Excellence in PhD Thesis. His research interests lie at the intersection of ethics and epistemology. He is an Associate editor of the Springer Journal – *AI & Society*. His publications have appeared in journals like *Philosophia*, *Journal of Applied Philosophy, Kriterion-Journal of Philosophy*, and *JICPR*.

Luke Roelofs is an Assistant Professor of philosophy at the University of Texas at Arlington, working on the metaphysics, epistemology, and ethics of consciousness. They are especially interested in thinking about agency and value at multiple scales, from human individuals both up to collectives and down to their subsystems. Their 2019 book, *Combining Minds*, explores the metaphysical foundations for composite minds, with special reference to panpsychist theories of consciousness; their latest book, *Empathic Reason (2026)*, tries to link the rationality of altruism to the need for mental simulation to properly understand other minds. When not doing philosophy, they enjoy learning about animals, playing board games, and hanging out with their cats.

Carol Rovane is Violin Family Professor of Philosophy at Columbia University. Her current research focuses on several interrelated topics: the first person, personal identity, relativism, the foundation of values and

group vs individual responsibility and some new problems for liberal theory. Her first book, *The Bounds of Agency: An Essay in Revisionary Metaphysics* (1997), challenges traditional views on personal identity and argues that personhood is defined by a commitment to rational agency rather than physical or psychological continuity. Her second book, *The Metaphysics and Ethics of Relativism* (2013), develops a sophisticated account of relativism that suggests that it has a consistent logical, metaphysical, and practical significance.

David P. Schweikard is currently a researcher with the ERC project "The Glue of Society: A Social Ontology of Social Cohesion" at Aarhus University in Denmark. He previously held positions at several German universities, including Düsseldorf, Flensburg, Münster, Cologne, and Duisburg-Essen. His published work is mainly in social and political philosophy, focusing on collective action, collective intentionality and responsibility, social ontology, political obligation, civil disobedience, and international political theory. He serves as Co-Editor-in-Chief of the *Journal of Social Ontology*.

Hans Bernhard Schmid is Professor of Political and Social Philosophy at the University of Vienna. He is a founder of the International Social Ontology Society, the founding Editor-in-Chief of the *Journal of Social Ontology*. Latest book: *We, Together – the Social Ontology of Us* (Oxford University Press 2023). Areas of Specialization: Social Ontology, Phenomenology, Philosophy of Social Science.

Manidipa Sen teaches philosophy at Jawaharlal Nehru University. She specializes in the areas of philosophy of mind, philosophy of language and epistemology with focus on questions concerning self-knowledge, self-Consciousness, and related issues in language. Her publications include *Thinking about the World: An Essay in de Re Thought and the Externalist-Internalist Debate* (2008) and several edited volumes. She has published articles in national and international journals.

Bill Wringe is an Associate Professor in Philosophy at Bilkent University, Ankara, Türkey, and Honorary Visiting Research Professor in the

Department of Philosophy at the University of York, UK. His interests include both social and political philosophy and the philosophy of mind, and he has a particular interest in collective obligations and in the philosophy of emotion. He is the author of *An Expressive Theory of Punishment* (Palgrave 2016) and of articles in a number of leading philosophy journals, and is the co-editor (with Säde Hormio) of *Collective Responsibility: Perspectives on Political Philosophy from Social Ontology* (Springer, 2025).

OceanofPDF.com

Acknowledgments

Edited volumes are like potluck dinners. They bring together a variety of cuisines, prepared in different places yet shared around one common table. Their value lies not just in the distinctiveness of individual preparations but also in the occasion they create for ideas to come together in conversation. If this gathering of essays has made for a worthwhile intellectual meal, the credit belongs entirely to the contributors who placed their trust in me and allowed their essays to become part of this dialogue. I am deeply grateful to each one of them.

The making of this collection, however, was itself the outcome of a cohesive endeavour. Many people contributed to bringing about this final product. Among these are the people who supported us in putting together an excellent workshop on social ontology in 2023, from which this volume eventually emerged.

I owe special thanks to Smita Sirker and other colleagues at the Centre for Philosophy, JNU for facilitating the workshop in countless ways. With their support, the workshop was able to bring together more than a hundred participants, including faculty members, independent scholars, doctoral researchers, and master's students from across universities and institutions. Each one of them contributed to the success of the event, and I thank them all for showing keen interest in this relatively new area of philosophy.

Four PhD students of mine, the ever-disciplined Astik, the resourceful Abhinav, the meticulous Aditi, and the perfectionist Harshita, were real pillars of strength. I also want to thank Shimran, another student of mine, for her assistance in the final stages of formatting the manuscript.

The International Social Ontology Society (Vienna) provided the financial support for the workshop, for which I express my sincere

gratitude. I appreciate the leadership of the Society for their goodwill and outreach.

I would also like to thank two beautiful minds, Olle Blomberg and David P. Schweikard, both contributing authors to this volume, whose support was crucial during the preparation of this manuscript. Their generosity came not only in the form of helping me imagine the structure of this work but also through their constant encouragement throughout the editorial process.

Lastly, but no less importantly, I thank my better half, Swargajyoti Gohain, whose academic companionship and everyday inspiration were instrumental to the completion of this project.

OceanofPDF.com

Introduction

Ontology of the Social: Agency, Attitudes, and Ethics

DOI: 10.4324/9781003612827-1

Ontological reflections on society are not new. Throughout history, philosophers from diverse traditions have examined the nature and reality of the social world, each distinguishing itself from others by deploying innovative styles and approaches. Within this broad tradition, social ontology—or, more appropriately, analytic social ontology—emerges as a field with a distinct commitment to investigating conceptual and metaphysical issues related to human society. As an area focused on the idea of the social, social ontology investigates a wide range of phenomena of the social world, analysing their nature, properties, and composition with the help of the latest methodological tools and techniques. At its core, it seeks to explain how the social world, with all its constitutive elements and structures, exists and persists over time with a distinct ontological status.

Traditionally, social entities such as groups, kinds, organisations, genders, or institutions have hardly been part of discussions in analytical metaphysics (Mason and Ritchie 2020). In most philosophical writings, where individual entities are treated as the primary objects of analysis, there is an underlying assumption that social entities are reducible to more fundamental entities such as individuals and their properties. This conviction is not only evident in the writings of individualists, including those in the Weberian tradition, but also of broader strands of reductionism that systematically privilege individual-level explanations over collective ones. This individualist orientation served, for a long time, as the *de facto* standard against which philosophers in the analytic tradition assessed competing theoretical frameworks.

Of late, however, a significant number of philosophers have called this individualist focus into question and have reimagined the scope of analytical inquiry about society. They contend that there is no justifiable reason why entities of the social world should receive less attention than their non-social counterparts.[1] They suggest that rather than dismissing these entities as non-fundamental, mind-dependent, or non-natural, they must be rigorously studied by paying adequate attention to their irreducible features. Over the past three decades, this reassessment of social entities has been accompanied by a parallel recognition that many concepts in orthodox analytic philosophy are themselves socially embedded. Many contemporary thinkers, including critical theorists, social constructivists, and feminist theorists, have highlighted the way in which social dimensions make an ineluctable entry into the conceptualisation of ideas such as knowledge, agency, autonomy, self, and consciousness. They raise crucial questions about the inherited individualistic biases and limitations in understanding these concepts.

Observing these developments, commentators have also begun to anticipate a new, even more radical turn in philosophy, which they call a ‘social turn’ in contrast to the linguistic turn that shaped much of twentieth-century analytic philosophy (Ásta 2024).[2] Whether such a turn is indeed taking place within mainstream philosophy remains a debatable issue—one that posterity may be better positioned to judge. What is less debatable, however, is the fact that analytical examinations of social entities are now much more acceptable than ever before. Analytic philosophical practices in recent years have increasingly stressed the need to engage with the concepts related to concrete social realities such as sex, gender, race, and disability without much reservation. As this intellectual approach gains traction increasingly, it demands critical attention, not only from professional philosophers but also from researchers and scholars in adjacent fields.

This volume is not an introductory text, nor is it a comprehensive inventory of topics falling under the rubric of social ontology. Instead, it is an attempt to examine the development of some of the core concerns that have shaped the field through sustained and critical engagement. The aim is to illuminate how particular conceptual issues, categories, and problems have been framed, refined, and contested, and to assess what is at stake in

their ongoing treatment. The themes covered in this volume are diverse yet overlapping; they are: collectivity, social groups, plural and collective actions, agency, we-attitudes, and collective responsibility. While these themes do not represent the entirety of social ontology, careful analyses of these phenomena and their interconnections can yield valuable insights into the field's current status and future direction.

An important feature of these contributions is their shared attempt to examine conceptual matters in ways that transcend the conventional boundaries of the philosophy of the social science. The discussions span multiple areas of specialisation, encompassing philosophy of language, mind and action as well as moral and political philosophy, game theory, and business ethics. Keeping the phenomenon of sociality at the centre, the authors adopt diverse strategies. Some explore previously unexplored or underexplored issues. Others advance or analyse established claims and categories. Still others offer new analyses of phenomena that are informed by various specialised philosophical sub-fields. Out of 13 essays, 9 were presented at the workshop 'Signs of the Social: Ethics and Ontology', held in December 2023 at Jawaharlal Nehru University, New Delhi, under the aegis of the International Social Ontology Society (ISOS). The remaining four chapters are new contributions by Seumas Miller, David Schweikard, Luke Roelofs, and Olle Blomberg.

1 Ontology of the Social: A Snapshot

Imagine a brain scientist born and raised in an artificial environment with no exposure to the outside world.[3] Despite her isolation, she has somehow acquired comprehensive theoretical knowledge of human psychology, much like Frank Jackson's neurophysiologist Mary, who knows everything about colour and colour perception without ever having had visual colour experiences (Jackson 1986). This scientist has mastered the knowledge of human psychological states, their structures, and the various conditions behind every human behaviour.

One day, she leaves her artificial home and interacts with people in the real world. In her first encounter, she sees three sets of people: a family enjoying ice cream at a nearby parlour, a group of young people making

reels for social media, and a crew of uniformed paramedics curiously observing her movements. Since she possesses complete knowledge of individual human psychology and behaviour, we might think that she could easily make sense of these social activities as effectively as we do. At first, she would probably require some guidance and nudges, comparable to a bedridden patient reintegrating into society after a prolonged hospitalisation. Still, we might assume, she would eventually be fine without any significant obstacles in grasping what is going on around her.

The question, however, is whether the brain scientist truly knows and sees the world as we do. In explaining the behaviour of the sets of people she meets, we are naturally inclined to appeal to concepts of social and institutional relationships, such as family membership, peer groups, and professional roles. After all, their activities are not merely collections of parallel individual behaviours. Rather, they are activities of people in specific relationships under particular contexts. Given this constitution of social reality, can the brain scientist legitimately claim to understand what it means for these people to exist within such institutional relationships? Does she comprehend what entities like family, friendship, or professional relationships actually signify to those involved?

The idea of introducing this thought experiment is not really to establish a thesis of ontological irreducibility of social entities. Rather, it is meant to bring to light an explanatory difficulty, i.e., how an agent with adequate knowledge of individual-level psychological facts may nevertheless fail to understand the social world in the way we ordinarily do. A proper ontological understanding of the social must be able to recognise this gap, even if it is committed to explaining the sociality in terms of individualistic or reductivist understanding.

The experiment, in effect, brings into focus two distinct but related limitations in the brain scientist's understanding of the social world. The first is an epistemic limitation. The scientist does not understand the social world in the way we do. Despite having complete psycho-physiological information of individuals, the scientist lacks something very crucial. What she lacks is an understanding of social meanings, norms, and roles. While she may be able to carefully observe everything encountered after release, such observation does not reveal the reality of things like family, social media, paramedic crews as groups, or even places like an ice cream parlour

as an institution. She cannot comprehend these as her objects of observation. To understand what a family is, what it means to participate in social media practices, or what it is to act as a paramedic, one must possess a standpoint internal to relevant social practices. The scientist can very well recognise the patterned behaviours, she radically lacks epistemic access to the meanings those behaviours have for those participants themselves. By internal standpoint, I do not mean phenomenological or first personal experience. Rather, I refer to it as having the ability to occupy a perspective that enables agents to recognise roles, actions, attitudes, and the like as shared rule-governed practices.

But this limitation is not restricted to epistemic limitations alone. There is also an ontological aspect to the scientist's failure to understand the social world. Social or collective entities such as families, professional crews, peers, and institutions are not just aggregates of individuals whose existence can be made sense of from their physical or psychological properties. They are more than that. They are constituted by individuals who possess certain shared concepts, intentionality, norms, rules, and the like. These features are available only to participants in relevant social practices, which the brain scientist is deprived of. So, what she encounters are individual behaviours, but not the behaviours of families, friends, or professional units as such.

Thus, although the brain scientist may know everything there is to know about those individuals she encounters after her release, she is still unable to recognise the social entities that these individuals collectively form. Her knowledge is the knowledge about persons considered in isolation. It is not a knowledge that recognises their identity and status in a given social and institutional context. What is missing is not just additional empirical information about these individual minds and bodies, but a crucial understanding of social meanings, roles, facts, and normative structure through which social reality or collectivity is formed. This gap suggests that social reality is not reducible to, nor fully derivable from, the psychological facts.

To understand social entities such as families, friendships, institutions, or professional and personal roles, one must occupy a standpoint internal to the relevant social practices (Haslanger 2012). Full knowledge of the social world requires more than exhaustive theoretical knowledge of individual

human psychology. It requires access to a form of social life that makes such entities intelligible and ontologically possible.

2 Social Ontology: Concepts and Contours

A number of concepts, themes, and debates have surfaced in discussions concerning sociality, both within social ontology and in adjacent areas. In this section, I briefly highlight a few whose relevance and influence resonate throughout all the chapters of this volume. The discussion does not aim to provide a comprehensive picture covering the breadth and complexity of these issues, nor does it seek to take a particular stance in defending or explaining them. My more modest aim is to simply offer the reader a preliminary orientation which may prove helpful while engaging with the chapters that follow.

Foremost among them is the idea of collectivity. This idea embodies two essential facts: the singularity of an entity and the multiplicity of its constituent elements. The difficulty philosophers face—whether in articulating a collective mental state or a moral phenomenon associated with a collective context, or in identifying the origins of relevant agency in a collective setting—is that of handling the combination of these two facts. For instance, with regard to the concept of collective intentionality, which is broadly understood as the underlying *mental mark* of, say, a collective action situation, philosophers continue to be divided over whether its collective essence is found in its mode of expression, the content of the state, or the subject who bears it. Likewise, philosophers engaged in debates about collective responsibility are concerned with whether the core of its being collective resides in the jointness of realising a collective outcome, or whether it is embedded in the unity and oneness of the group that does the realising, or whether it is constitutive of the way the responsibility is distributed among the relevant individuals. The crux of this debate lies in the fact that there is no one single way of understanding what collectivity involves, nor of understanding how it collectively shapes our understanding of various social and communal phenomena. We find all these positions represented in almost all of the chapters in this volume, with a recognisable pattern of thought appearing in most of them.

A related but distinct theme is the notion of social groups, their types, and classifications. Our social space is filled with a wide variety of social groups, from cultural and ethnic groups to families, organisations, institutions, nations and states, to conglomerations, committees, and teams. But what functions, if anything, as the unifying factor among these varied groups, such that we can categorise them all as social groups? Some groups have formal structures, others lack them. And yet others persist without displaying any discernible pattern. Some are static and enduring, while others are fleeting and short-lived. And yet others are dynamic with a capacity to adapt form and function to changing circumstances. Such variation at the level of duration, structure, and cohesiveness presents significant challenges to any comprehensive theory of social ontology. However, this does not mean that this diversity has deterred philosophers from attempting to analyse the essence of grouphood or collectivity. Contributors to this volume work on a range of groups, with different structures and scopes in their pursuit of understanding human social phenomena. Each contribution confirms the insight that social ontological discussions must attend to and respect the uniqueness of different kinds of social groups.

The tension between individualism and collectivism, the two opposing schools of thought for comprehending social facts, entities, and attitudes, is another recurring theme in most social ontological discussions.[4] These doctrinal positions differ, among other matters, on how thick or thin the appropriate notion of collectivity should be in our analysis of the social world. Their differences, roughly speaking, may manifest at two distinct levels: the ontological and methodological levels. At the ontological level, collectivism maintains that social realities are fundamentally grounded in the reality of group entities. At the methodological level, it suggests that explanations of social realities or social phenomena cannot be entirely based on individuals and their interrelations. Individualism, conversely, holds that, at the ontological level, there are no non-reducible social entities- only individuals can have an existence of their own. At the methodological level, individualism holds that social phenomena can and should be explained by demonstrating how they emerge from individual

actions, which in turn must be explicated with reference to the properties of individuals.

Both individualism and collectivism admit the possibility of multiple layers of explanations depending upon how their core assumptions are articulated and defended with regard to a particular social phenomenon. While collectivism of a robust form does not enjoy a respectable position in most quarters, many philosophers nevertheless find it extremely difficult to dispense with collectivism altogether. After all, at the rudimentary level, collectivism is ineluctably embedded in our everyday vocabulary. One might suggest that there exists a class of appropriate linguistic expressions typically reserved for collective entities such as ‘winning a football game’, ‘defeating a country’, or ‘delivering a court judgment’ (Will Kymlicka 1989). None of these expressions bears any meaning when applied at the level of individuals, but each becomes intelligible when we use them in the context of social entities. This recognition shows that collectivism is neither a new thesis nor an alien method of approaching the reality question of sociality.

Individualism, the arch-rival of collectivism, also comes in a wide variety of forms. It ranges from an atomistic interpretation to a more sophisticated articulation that goes beyond the conventional ontological and methodological commitments. Even though atomism as an ontological thesis has failed to impress most social theorists, reductivism as its core explanatory approach is still dominantly present in many accounts. It plays a crucial role in a number of theories that consider formal individualism as the only way forward. According to this form of individualism, groups and collectives, unlike what a strong collectivist would hold, are not the kinds of entities that hover over and above their constituent individuals. Instead, they emerge from the dynamic interrelations of individuals with respect to a particular fact or phenomenon. This frame of analysis gives us a more refined explanation of how social reality operates at multiple levels simultaneously.

Another important concept that has attracted our attention in these discussions is agency. If social entities are indeed irreducibly real and are capable of demonstrating capacities analogous to those of individual agents, what kind of agency do they possess? Some philosophers argue that this agency question cannot be satisfactorily settled at the individual level. They

maintain that collective entities can very well be agents in their own right, given that, on functional grounds, such entities' beliefs and preferences cannot be reduced to the beliefs and preferences of individual constituent agents. Such a group agent can have its own agential status and autonomy, distinct from those of its constituent individuals. This thought has an illustrious history. Thomas Hobbes, for instance, in his *Leviathan* (1994, Chapter 16 of his (1651) magnum opus), notably argued that a group agent can come into existence when a collection of people authorises a voice on a particular matter to go by that. But how should we understand the nature of this 'voice' that allegedly works over and above the individuals? To qualify as a genuine agent, a collective entity must exhibit a degree of independence from its environment in which it operates and its constituent members who make it what it is. It must function as a unified entity rather than as a mere metaphorical way of speaking about the actions of its constituent individuals. This distinction matters profoundly for understanding whether group agency represents something genuinely novel or merely an aggregation of individual actions.

In recent times, several important accounts of group agency have been developed. Christian List and Philip Pettit (2011) maintains that states, nations, committees, and corporations fulfil the basic condition of being an agent when their individual members are committed to go by a group-level decision-making process which allows them to arrive at group beliefs and group intentions (in the functional sense) whose contents are distinct from individual members' individual beliefs and intentions. Similarly, Carol Rovane (1997) thinks that individual mental states together can give rise to a *rational point of view*—a bundle of interconnected mental states from which a collective entity can act and operate. While this picture of group agency has attracted a lot of attention in recent social ontological discussions, another picture of a similar phenomenon (but not the same) has also been under serious consideration, in which the rationale of agency is not overtly individuated. It begins by asking: What is it to act in a situation over which one does not have complete control, yet in which one is not agentially irrelevant? Or, how does one act in concert with others while pursuing something that matters to them collectively? There are countless social phenomena that fall under levels such as shared acts, joint actions, and collaborative activities. One way to highlight the importance of the

agential angle of such social phenomena is to ask a question, in a Wittgensteinian spirit: What happens when we subtract what each of us does from what we do together?[5] Is there any distinction between actions of individuals that together constitute our shared activity, and the actions that they perform parallelly in a given time and space? John Searle (1995) thinks that the answer lies in the intentional makeup of these actions: the former grounded in We-intentions, the latter merely consisting of individual I-intentions. We-intentions, in this sense, are the intentions that explain the possibility of genuine collective actions constituted by actions performed by individuals. Many writers, however, find such explanations unsatisfactory, given that there are social phenomena that resist analysis based on collective intentionality. Consider global climate change caused by anthropogenic carbon emissions—a collective phenomenon that lacks intentional makeup among constituting individuals to cause climate change, yet it is something that arises from the cumulative individual contributions, for which there are important moral implications.

Can we understand the metaphysical reality of a social phenomenon merely on the basis of its moral implications? For instance, collective responsibility is essentially a practical implication whose legitimacy depends on the metaphysical possibility of collective action, agency, or intentionality. To understand the rationale of moral responsibility as a practical implication, we must first understand what it is and where it comes from. Taking a contrary position, however, one might put the justificatory arrow the other way around—following a strategy adopted by P. F. Strawson (1962) while dealing with the philosophical deadlock between free will and determinism. The possibility of collective responsibility, one might argue, does not necessarily require robust singular agential involvement; neither does it require any cohesive intentional structure or design behind every morally condemnable collective outcome. What it requires instead is that there exists a crucial moral fact for which there has to be moral responsibility. Owing to this intuition, many might even doubt if we are at all required to follow the same theoretical template of individual cases at the collective level, i.e., treating the individuals as the paradigm case and then mechanically deploying the resulting conditions and categories to collective entities.

Collective responsibility remains a deeply contested issue among philosophers and ethicists alike. While disagreements could go in multiple directions, the most crucial one concerns the issue of attributing it in concrete terms. Some believe that collective responsibility is just a shorthand for referring to the responsibility of constituent individual members (Miller), and others consider it as something that refers to a special entity, 'the collective', of which the individuals are parts (Roelofs, Neog, Rovane, and Collins). While neither seems to be fully justified in dismissing what the other has to say, both build their own coherent theoretical fortresses. The former frames it as individuals' responsibility in collective contexts, and the latter frames it as the responsibility *of* the collective in question. This distinction has significant ontological implications. The latter position adopts a realist stance in that it treats collective responsibility as an irreducible phenomenon. The former rejects this view on grounds of ontological parsimony and suggests that individualism is the only acceptable meta-theory of morality and responsibility (Jan Narveson 2002). Both these positions offer benefits, owing to the avenues they open for the explanations of different forms of harm and wrongdoing perpetrated by groups of people. Several chapters of this volume highlight the benefits and burdens of these two approaches while attempting to make sense of the social nature of morals and metaphysics.

3 Chapters in Brief

Now, the remainder of this introduction presents the basic thrust of each chapter, keeping in mind their unique contributions to the volume. David Schweikard, in the inaugural chapter, investigates an apparent conflict between social constructionism and metaphysical realism, arguing that something can be socially constructed and yet still be real. He begins by identifying what he calls a *metaphysical bluff*—a strategy in which ideas of mind-independence are used to shield certain social practices from criticism. Drawing on his earlier work with Arto Laitinen, Schweikard then introduces a tripartite framework consisting of cognitivism, success theory, and objectivism. Departing from the traditional realism which typically

maintains that something must exist independently of the mind in order to be real, he defends a more minimalist version according to which social facts, though practice-dependent, possess genuine causal and normative efficacy. He further argues that this shift in perspective is necessary for understanding social change. Because social reality is grounded in collective activity, it can also be changed through collective action. Recognising oppressive social structures as real is thus an important condition for changing them.

Social realities are often realised through institutions. These institutions, particularly those with established structures of their own, are capable of functioning because of the normativity of various kinds. Seumas Miller, in his paper, discusses the moral normativity of social institutions and analyses its three sources—needs, joint productive activity, and reciprocity. In addressing these, Miller assumes that social institutions are organisations. For him, unlike desires, needs are objective, limited, and generate moral obligations because unmet needs cause harm. When aggregate needs exist within populations—for food, healthcare, security, education—they create collective moral responsibilities that institutions are required to fulfil. Organisations with a professional outlook often exemplify this; for instance, doctors address health needs, lawyers serve justice needs, and fire-fighters perform public safety needs. Based on Locke's theory, Miller considers joint productive activity as the second source that acquires moral rights to those goods or appropriate rewards. He notes that only value-producing labour generates such rights, specifically labour that creates goods satisfying genuine needs. Reciprocity, the third source, refers to the intra-institutional and inter-institutional reciprocal moral rights generated by members who are producers/consumers within a complex. In the former, reciprocal relationships exist between producers and consumers of collective goods. But in the latter, complex webs of reciprocal moral rights emerge as individuals specialise in producing one collective good while consuming multiple others across society's institutional network.

Bernhard Schmid addresses an intriguing controversy found in the philosophy of language that concerns the role of uptake in illocution. The controversy asks whether a successful communication requires that the addressee 'take up' the speaker's intended meaning, or whether the speaker can successfully perform acts like commanding or warning, even when the

addressee fails to recognise it. Positioning himself against the classical speech act theory held by J. L. Austin and Adolf Reinach, which considers uptake as something a speaker must ‘secure’ from addressees, Schmid argues that this view is often marred by a speaker’s bias that treats the audience as a mere target. To address this, Schmid proposes that illocutionary acts should be understood as plural actions—joint activities in which both speech and uptake are active contributions that require shared intention among participants. This means successful communication depends on both parties working together through discussion and negotiation, rather than one side controlling the other. Schmid also engages with some of the latest experimental works by Fisher, Francis, and Townsend (2023). These works suggest that, contrary to what the classical theories suggest, laypeople often believe that speech acts, especially refusals and warnings, occur even when the audience fails to grasp the speaker’s intention. He, however, interprets empirical results not as an outright rejection of the uptake condition, but rather as an instance of participants’ involvement in a cooperative social act where the meaning of words depends on the intentions and normative expectations of the participants. Schmid finally argues that the determination of what a speech act counts as is dependent on a jointly negotiated consensus between all the participants in the act.

Building on her previous work, Carol Rovane develops a distinctive account of group agency that departs from standard social-ontological approaches towards it. While most theories of social ontology treat group agency as a social phenomenon, Rovane’s cases of group agency are not social phenomena. This consideration is based on a normative and revisionary account of personal identity, which challenges the traditional assumption that one person always equals one human animal. Rovane argues that personal identity is not a metaphysical given, but rather a self-constitution through a commitment to rational unity in thought and action. This normative account reflects a version of Locke’s distinction between personal and animal identity, which allows us for the existence of group agents where multiple humans function as a single person and multiple persons existing within a single human life. Extending this, Rovane aims to rearticulate moral responsibility as forward-looking engagement rooted in normative commitments rather than past actions. Drawing on Strawson’s

‘participant stance’, she emphasises the need to redefine responsibility as a matter of a forward-looking engagement with these unique agential points of view. Through ‘hybrid cases’ such as the Philosophy Department and the Manhattan Project, she shows how group agents can be formed across multiple human lives, thereby effectively challenging the common assumption that human beings are ‘natural persons’ with fixed boundaries of responsibility.

Bill Wringe addresses an important challenge to the claim that non-agential groups (such as humanity as a whole) can bear moral obligations. Wringe thinks that many arguments for the existence of non-distributive backward-looking collective responsibility appeal to ‘responsibility gaps’. A responsibility gap is a situation where it seems as though some entity or entities must be responsible for a bad outcome, even though no individual is. Arguments of this sort often seem to make a mistake. They assume, implicitly, that when a bad outcome occurs, there must be a fixed amount of blame to be distributed. Philip Pettit, David Copp, and Sara Rachel Chant, for instance, assume that we can identify an outcome as one for which some entity or other bears responsibility without establishing which that entity is. According to Wringe, we can see that this is a mistake by thinking about the functions of blame. These functions tell us that there is no good reason for thinking there must be a fixed amount of blame to distribute. However, there are significant differences between the purposes of assigning blame for bad outcomes in the past and assigning responsibility for possible bad outcomes in the future. So arguments for the existence of non-distributive forward-looking responsibility are not vulnerable to the same kind of criticism.

Random collectives are standardly viewed as mere aggregates of individuals. Most philosophers argue that their alleged moral standing can and ought to be determined by the actions or omissions of individuals who happen to share certain traits in a given context. Neog challenges this widely held position by asking, what if we, instead of focusing on whether random collectives possess unified agency, consider whether a collection of relevant inactions in a particular situation itself constitutes a morally important phenomenon. Exploring the implications of this question, Neog takes up an unorthodox case to show how a small-scale, random collective can be understood as an appropriate target of non-distributive moral

responsibility by focusing on its moral success rather than its moral failure. He argues that when the discharge of an obligation is understood on representational grounds, the collective may itself become a proper subject of moral assessment.

Stephanie Collins takes up an important social ontological issue that raises the question about the legitimacy of corporate entities. She asks whether the concept of legitimacy, conventionally used exclusively in the context of nation-states, can be meaningfully extended to business corporations. The motivation for her inquiry stems from contemporary debates about corporate power, particularly concerning large technology companies like Amazon, Google, and Meta. While such entities are growingly described as facing a ‘crisis of legitimacy’ due to their roles in democratic polarisation, privacy violations, and other harmful outcomes, it remains unclear whether legitimacy can appropriately apply to business entities. Collins begins by outlining a theory of collective agency that is applicable to both nation-states and business corporations. She then develops a broad and ecumenical account of legitimacy that is particularly applicable to business corporations. A business corporation is legitimate when it adequately performs its function of producing certain characteristic goods via the distinctively corporate structure. At a minimum, a legitimate corporation must respect basic rights and be responsive to the needs of stakeholders, within the domain of the goods it characteristically produces.

Attributions of moral responsibility for large-scale collective actions are rather difficult. Its difficulty arises on many fronts, including the fact that it involves the entanglement of many people in different ways to different extents for different reasons. Luke Roelofs focuses on one specific aspect of this difficulty, involving the attempt to identify the agent of a collective action. He calls this the ‘problem of overlapping collectives’: For any group we might identify as the agent responsible for the act, there are slightly smaller or slightly larger groups that seem just as plausible. He argues that this challenge can be met if we extend the approach known as ‘manyism’ from individual objects to social groups, though doing so brings into sharp relief certain differences between human individuals (fairly well-individuated) and social groups (often very loosely individuated). These differences make collective agency messier and harder to describe, but do not by themselves undermine its agential character.

Jeremy Koons, in his essay on team reasoning, considers the possibility of a variety that goes well with a deontological framework. He argues that prevailing game-theoretic approaches to morality fail to capture essential features of deontological obligations. The current game-theoretic literature distinguishes between different kinds of coordination games, such as pure coordination games (choosing which side of the road to drive on) and assurance games (stag hunt, where the higher payoff is achieved if both players cooperate). These approaches, however, almost uniformly rely on consequentialist presuppositions, which entail that morality, if justified in game-theoretic terms, must take the form of an assurance game. This poses a serious challenge to interpreting moral obligations on deontological terms. Morality in a deontological framework invariably requires that agents ought to perform actions even in situations when there is no assurance that others will comply. If morality is understood entirely as an assurance game, such unilateral compliance would appear unjustified. Drawing on the work of Wilfrid Sellars, Koons argues that team reasoning need not be understood in consequentialist terms. While all team reasoning involves an element of coordination, our moral practices largely do *not* serve a coordinating function and cannot be understood as an assurance game. It can be rational, and even obligatory, to act non-cooperatively per the dictates of morality. We can thus articulate a deontological form of team reasoning, Principled Team Reasoning.

Reflective equilibrium is a rational method for developing coherent moral and political beliefs by balancing general principles with particular judgements. Sushruth Ravish discusses it in the context of understanding the phenomenon of moral progress. While Reflective Equilibrium (RE) has been widely accepted as a premier method of moral inquiry, its conventional approach fails to explain how moral progress occurs through collective deliberation and argumentation. Ravish argues that a suitably amended version, called social RE, has the potential to fulfil that promise. Drawing on the literature on collective epistemology and social interactionist theories of reasoning, he argues that certain groups can overcome individual cognitive limitations through appropriate deliberative processes. To Ravish, the case of the Indian Constituent Assembly Debates, which took place between 1946 to 1949, provides an empirical validation. It

shows how the Assembly's unprecedented achievement of universal adult suffrage exemplifies successful collective moral reasoning.

Joint activities pervade our daily life. One of the foundational features of joint activities, according to Olle Blomberg, is that each participant of such activities is rationalised and guided by a 'we-intention'. A joint activity J is a joint intentional activity if and only if it is appropriately caused and coordinated by the participants' we-intentions to J, where these we-intentions consist of ordinary mental states with collective contents. Drawing on Michael Bratman's work, Blomberg argues for two minimally sufficient conditions for a participant to have a we-intention to J: First, she should have an intention that she and the others J, by way of each of their intentions that they J together, and by way of co-realisable sub-plans (this is Bratman's so-called 'intention condition'). Secondly, she should have a belief or assumption that the end that they J together is shared, such that they each intend it. Blomberg calls this the Bratman-lite account of we-intention. He then addresses certain objections which suggest that such a content account may not be able to explain how we-intentions can arise from 'team reasoning', which Koons' essay deals with, a form of practical reasoning that sometimes leads individuals to engage in joint intentional activity. Blomberg argues that these objections do not succeed and that the content approach to we-intentions is fully compatible with theories of team reasoning.

Consciousness is understood as something that typically belongs to individual minds. So, one wonders whether individual consciousness, understood from the first-personal singular perspective, gives rise to a genuinely first-personal plural perspective of collective consciousness. Bijoy Boruah addresses this foundational problem in his essay by making us aware of the intuitive primacy of individualist ontology. Boruah confronts that difficulty by denying two obvious accounts: the immanentist account that posits we-sociality as an innate capacity and the aggregative theories that fail to generate genuine collective unity. He instead argues for an emergentist interpretation in which we-consciousness develops through a mediated process of intersubjective engagement. Using the insights from phenomenological writings of Edmund Husserl, Martin Heidegger, Martin Buber, and contemporary interpretations of Dan Zahavi and Bernhard Schmid, Boruah draws a distinction between what he calls 'thin' we-

sociality (minimal dyadic I-Thou relations) and ‘thick’ we-sociality (multitudinous collective identity). He argues that the second-person perspective—the I-Thou relationship characterised by dialogical reciprocity and ‘ontic directness’—serves as the structural mediating link between first-person singular and first-person plural consciousness. *We*-sociality thus, according to him, arises from a second-personal engagement which allows individuals to transcend their singular subjectivity while preserving their individuality within the collective whole.

Self-knowledge, put simply, refers to a kind of knowledge that one has about one’s own mental states. Collective intentionality, a crucial thesis of social ontology, presupposes a form of self-knowledge in which participation in collective intentional states requires that agents know they are the subjects of those states. Manidipa Sen critically analyses this relationship by focusing on a few key dimensions of self-knowledge. Sen argues that the awareness of oneself as a singular ‘I’ is the necessary foundation for the very concept of the ‘we’. In this context, she critiques a well-known view, the ‘community first’ view, which suggests that our sense of self is merely a social construct derived from our participation in a linguistic or normative community. Subscribers to this view, philosophers like Lynne Rudder Baker and Dan Zahavi maintain that individual identity is socially constructed, in which the robust first-person perspective emerges only through social interaction within the normative and linguistic spaces of communities. Sen critically evaluates this ‘community first view’ by highlighting several difficulties, such as the distinction between sameness and identity, linguistic immunity to error, and others.

Notes

1. It is important to identify the constituent parts of a social entity as non-social entities rather than individual entities. While social entities are often contrasted with individual entities, this contrast can at times be misleading, since social entities can themselves be described as individual entities under some description. To avoid this ambiguity, it is therefore preferable

to characterise the constituent parts of social entities as non-social entities. I am indebted to Olle Blomberg for drawing my attention to this.⏎

2. Ásta (2024) refers to a piece written by Kevin Richardson (2023) on the widely followed philosophy blog *Daily Nous*. Richardson claims that the social turn has arrived, and makes the following interesting remark: 'The linguistic turn is over. We partied hard, got hungover, and now we're trying to live as respectable adults… Today, a new revolution is brewing. Analytic philosophy is in the midst of a *social* turn.'⏎
3. A similar, but less developed, thought experiment is found in Deborah Tollefsen (2014).⏎
4. I do not attempt to introduce a fine-grained distinction between these two positions, nor do I provide a complete map of their complex of interconnections. Instead, my aim is only to present their theoretical sides in simpler terms and to highlight how debates and disagreements surrounding these two approaches have enriched social ontological discussions.⏎
5. Wittgenstein (1953: *Philosophical Investigations* $$ 621) famously asked a question that tremendously helps us in clarifying the distinction between actions and events. He asked, 'What is left over if I subtract the fact that my arm goes up from the fact that I raised my arm?'⏎

References

Ásta. 2024. 'Social Ontology: Where Now?', *Journal of Social Ontology*, 10(3): 7–10.⏎

Fisher, Sarah A., Francis, Kathryn B. and Townsend, Leo. 2023. 'An Empirical Investigation of Intuitions about Uptake', *Inquiry*: 1–33.⏎

Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*, Oxford: Oxford University Press.⏎
Hobbes, Thomas. 1994. *Leviathan*. Edited by E. Curley. Indianapolis: Hackett.⏎
Jackson, Frank. 1986. ‘What Mary Didn’t Know’, *The Journal of Philosophy*, 83(5): 291–295.⏎
Kymlicka, Will. 1989. *Liberalism, Community and Culture*, Oxford: Clarendon Press.⏎
List, Christian and Pettit, Philip. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford: Oxford University Press.⏎
Mason, R. and Ritchie, Katherine. 2020. ‘Social Ontology’, in Bliss, R. and Miller, J. T. M. (eds.), *The Routledge Handbook of Metametaphysics*, *pp. 312–324*. London: Routledge.⏎
Narveson, Jan. 2002. ‘Collective Responsibility’, *Journal of Ethics*, 6: 179–198.⏎
Richardson, Kevin. 2023. ‘The Social Turn in Analytic Philosophy: Promises and Perils’, *Daily Nous*, 8 August. Available at: (Accessed on 15th Jan, 2026)⏎
Rovane, Carol. 1997. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.⏎
Searle, John R. 1995. *The Construction of Social Reality*. Oxford: Oxford University Press.⏎
Strawson, Peter F. 1962. ‘Freedom and Resentment’, *Proceedings of the British Academy*, 48: 1–25.⏎

Tollefsen, Deborah. 2014. ‘Social Ontology’, in Cartwright, N. and Montuschi, E. (eds.), *Philosophy of Social Sciences: A New Introduction*, pp. 85–101. Oxford: Oxford University Press.⏎

Wittgenstein, Ludwig. 2009. *Philosophical Investigations*. 4th edn. Translated by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte. Oxford: Wiley-Blackwell. (Originally published 1953).⏎

OceanofPDF.com

1 Realism, Constructionism, and Social Change

David P. Schweikard

DOI: 10.4324/9781003612827-2

1 Introduction

The drawing *Bedtime* by the British cartoonist Will McPhail (2022) depicts a parent kneeling before their young child. The child, dressed in a nightgown, listens reluctantly to the parent saying: ‘Listen to me. Gender is a construct, society is a construct, money is a construct. But bedtime is very, very real’. The comedic value of this scene lies in a tension within the parent’s statement. While it may be straightforward to introduce a toddler to a thoroughgoing constructionist worldview that treats social structures and institutions as human inventions or at least as products of human activity, the parent eventually reaches a limit. Bedtime is presented as the strict limit where the idiom of ‘constructs’ fails to apply.

For the adult in this story, the middle sentence likely appears as a series of truisms. They are perfectly comfortable using the jargon of ‘constructs’ when discussing gender, society, or the institution of money. However, the child is meant to learn that not everything is subject to human construction. Against the suggestion that social kinds are open to negotiation or alteration, the parent insists that bedtime exists beyond the scope of human making and deems it ‘very, very real’. This presents the toddler with a distinction between what is practice-dependent and what is mind-independent, with the latter serving as the traditional, heavy-duty benchmark for what is to be considered real. The child is thus meant to treat bedtime as unlike money or gender insofar as it is presented as non-negotiable.

Another source of the tension illustrated here stems from an ambiguity in the notion of a ‘construct’. In ordinary discourse, labeling something as constructed often suggests arbitrariness, dispensability, or a lack of constraint. Within social ontology, however, the term is typically used in the following, far more neutral sense: to say that a phenomenon is socially constructed is to advance a thesis about its dependence conditions, not to deny its existence or efficacy. In Ásta’s (2018) ‘conferralist’ framework, social construction is the active process by which our social environment bestows categories upon us. Categories that are constructed in this way are real in virtue of the way they structure social constraints and enablements. The failure to keep these senses apart encourages the inference from construction to unreality, an inference I reject in this chapter.

It will be useful to start by distinguishing between three claims that are frequently run together. First, the claim that a given social kind is not a natural kind. Second, the claim that it depends on human practices or attitudes. Third, the claim that it is not real. The first two claims are widely accepted within contemporary social ontology. The third does not follow from them. Much recent work attempts to articulate a conception of realism that accommodates the first two claims without entailing anti-realism (most prominently, Haslanger 2012).

Clarifying this helps to situate the present argument within a broader debate about realism. The issue is not whether social facts are constructed, but how construction relates to reality. If construction is treated as incompatible with reality, then social ontology must adopt either an eliminativist or fictionalist position, or some other form of anti-realism. If construction is instead understood as specifying a mode of existence, a realist account can accommodate both the dependence and the efficacy of social facts.

The parent’s move trades on an equivocation about reality. By calling bedtime ‘real’, they are attempting to exploit the perceived mind-independence of the physical world to shield off a negotiable social practice from revision. The cartoon illustrates the issues addressed below, though I will argue against the parent’s position. The first issue concerns the meta-debate over what warrants labeling a position as ‘realist’ – a second-order question that is largely independent of substantive, first-order questions about the nature of specific social entities like group agents or social

institutions. As Sally Haslanger (2012, 5) has noted, social structures and categories often seem ‘all too real’ precisely because they are causally and normatively effective. They shape patterns of behavior and social expectation, cause events, and generate obligations that constrain our actions. However, these structures and categories seem equally obviously dependent on human minds and social practices, creating a persistent tension for those who believe that reality must be strictly tied to mind-independence.

If we follow the traditional view of metaphysical realism, then anything mind-dependent, including anything socially constructed, is, by definition, not real. In this chapter, I argue that there is good reason to reject this requirement and instead adopt a cognitivist success theory regarding (statements about) social facts, treating gender, society, and money as real social phenomena. This requires abandoning the assumption that realism entails mind-independence. By treating constructionism not as the negation of the reality of social facts but as a claim about the specific mode of existence of social facts – namely, their practice-dependence – we can maintain that these facts are real without pretending they are ‘joints in nature’ (Barnes 2017, 2423). Finally, I argue that a plausible realism about social facts allows us to account for the possibility of social change. Because social facts are grounded in collective practices, they are in principle mutable rather than fixed. Constructionism should be understood as a claim about the practice-dependence of social facts rather than as a denial of their reality. I will conclude that the mind-dependent status of social facts is not a reason to doubt their reality. On the contrary, their mind-dependence is what makes them susceptible to transformation.

2 Are Social Facts Real?

The cartoon illustrates a distinction between constructs and reality that recurs in social ontology. The parent in the drawing refers to a variety of social phenomena, among them the social kind gender and the social institution of money. By listing gender and money as constructs, the parent suggests that these phenomena are fundamentally different from what is real, among which they count bedtime. In philosophical terms, this points to

a distinction between what we might call first-order metaphysical questions and second-order questions about what it actually means to be a realist about such entities (see Laitinen/Schweikard 2025). Social ontology primarily studies first-order metaphysical questions, inquiring into the nature and existence of group agents, organizations such as the United Nations, or social kinds related to race and disability. At this first-order level, it is also conceptually helpful to distinguish between social kinds and social properties. As Vincent Tanzil (2025) points out, individuals might belong to a recognized social kind (such as 'redheads') based on shared physical or biological traits, without necessarily sharing a distinct, underlying *social property*. These substantive debates target issues such as whether social entities are grounded in something more fundamental or are perhaps ultimately eliminable (see also Epstein 2018).

The meta-debate, which is my focus here, concerns the second-order question of what counts as realism in social ontology and on what grounds a specific view should be labeled 'realist'. Two authors may agree on a first-order answer, for instance, that gender is mind-dependent, while disagreeing over whether this commitment marks their view as realist or anti-realist. The parent's view in the cartoon implies that for something to be real, it cannot be a 'construct', but must be independent of our social practices and conceptions, and specifically exempt from negotiation or alteration. This reflects a widely held understanding of metaphysical realism according to which only mind-independent entities are considered real. Charles Mills notes that, with respect to a core debate in social ontology, a racial realist is traditionally someone 'who thinks it is objectively the case, independent of human belief, that there are natural human races; in other words, that races are natural kinds' (Mills 1998, 46). Under this restrictive definition, socially constructed entities and properties are not real, and social constructionism is categorized as an anti-realist view. In an early contribution to this debate, Michael Root (2000) argues against this assumption and shows that how we divide the world through our social practices creates categories that are deeply real in their consequences, even if they completely lack a natural or biological foundation.

In any case, realism in social ontology must reconcile a tension between our intuitive justification for treating social structures as real and their

dependence on humans' attitudes, activities, and practices. A plausible realism about social facts need not treat as real only those entities and facts that are mind-independent. Indeed, incorporating social entities into metaphysics requires revising restrictive assumptions about realism (see Mason/Ritchie 2020). Drawing on earlier work (see Laitinen/Schweikard 2025, for similar taxonomies cf. Sayre-McCord 1986, Haslanger 2012, and Mallon 2016), I demonstrate this by mapping different kinds of realism and anti-realism through a sequence of three guiding questions.

The first question (Q1) is whether a statement in the relevant social domain has a truth value at all. Examples of the statements referred to in this and the following guiding questions could be 'Megan is team captain', 'the meeting is at noon', 'Charlie is a woman', or 'the university's grounds end at this fence'. The relevant domain must be specified. One might, for instance, be a success theorist about institutional roles while remaining an error theorist about race. To answer (Q1) affirmatively is to adopt *cognitivism*, whereas to say such statements are neither true nor false, i.e. that they don't have a truth value, amounts to adopting *non-cognitivism* or *instrumentalism*. In philosophy of science, the latter view holds that theoretical concepts are mere tools for organizing observations and do not imply ontological commitment. They are seen as 'useful fictions' which enable us to systematize and predict observations without claiming that the underlying entities are real. For a cognitivist, however, such statements are truth-apt and aim to describe reality. This yields a *most minimal realism* (R1) according to which statements of the relevant kind are truth-apt. The view corresponds to what Uskali Mäki (2008) calls a 'realist attitude', which gives real existence and objective truth a chance, but is prepared to conclude that an entity referred to does not exist after all. One might, for instance, be a cognitivist about 'the divine right of kings' while remaining an error theorist about its truth.

The second question (Q2) asks: 'Is any statement in the relevant domain true?' Answering this one in the affirmative marks the transition to a *success theory*. While an error theorist, by contrast, might agree that social statements have truth values, they would argue that such statements are never true because the world contains nothing in virtue of which they would be true. An eliminativist about race, for instance, might construe the respective claims in a cognitivist fashion but maintain that there are no

races, meaning all such statements are false since the terms do not refer to anything. Realism in the sense of success theory, or *minimal realism* (R2), requires that some of these statements are literally true.

The third question (Q3) reads as follows: 'Are true statements in the relevant domain true in virtue of mind-independent facts?' This is the traditional benchmark for *objectivist realism* (R3), the view resulting from an affirmative answer to (Q3). If one assumes that only mind-independent things are real, then we must be anti-realists about, for instance, gender and money, for their existence rather obviously depends on social practices. But we can answer 'no' to this question while still saying 'yes' to the first two, thereby maintaining a form of realism that is appropriate for the social domain. With Geoffrey Sayre-McCord we should then say that realism is 'not solely the prerogative of objectivists' (Sayre-McCord 1986, 12), and variants like intersubjectivism – where statements about social facts are true in virtue of shared social practices – can be quite clearly realist. By distinguishing these questions, we see that the ontological status of social facts does not have to depend on their being mind-independent. On this account, statements about social facts are truth-apt, some of them are true, and their truth depends on intersubjectively sustained features of shared social practice.

The insistence on mind-independence as a condition on realism can be understood as a residue of a particular metaphysical picture, according to which the paradigmatic instances of reality are drawn from the natural sciences. On this picture, reality is identified with what would exist independently of any observers, any intentional activity, or any collectively sustained practices. However, extending this conception to the social domain is problematic. Social facts are not merely observed; they are constituted, at least in part, by the very practices through which they are recognized.

This does not entail any form of subjectivism. The dependence in question is not on the attitudes of particular individuals, but on structured patterns of collective activity. These facts are therefore not optional in any sense relevant to their reality. They impose constraints, generate expectations, and enter into explanations in ways that are characteristic of genuine features of the world.

The framework distinguishes semantic, ontological, and metaphysical questions that are often run together. In particular, it allows us to separate questions concerning semantic commitment from questions concerning successful reference, implying ontological commitment, and from questions concerning metaphysical grounding. Much of the disagreement in the literature can be traced to a failure to keep these dimensions distinct.

Consider, for example, the position of an error theorist about a given social domain. Such a theorist may accept that statements in that domain are truth-apt, thereby answering (Q1) in the affirmative, while denying that any such statements are true, thereby answering (Q2) in the negative, thereby subscribing to cognitivism, but negating realism in the success-theoretic sense. By contrast, an instrumentalist may reject (Q1) altogether, treating the relevant discourse as a useful fiction. These positions fall under the label of anti-realism, but they differ significantly.

The framework also clarifies the status of disputes about reference. An eliminativist may argue that certain social terms fail to refer, and hence that statements containing them are systematically false. However, this claim presupposes a standard of meaning. As Sally Haslanger (2005) has demonstrated regarding the semantics of social kinds, if we adopt an externalist approach to meaning, we find that our social terms often successfully track highly real, politically significant social categories even when our everyday, intuitive definitions of those terms fall short. Once this assumption is explicit, it can be evaluated directly rather than treating it as implicit in the very notion of realism.

One can be a realist by rejecting eliminativism and embracing a success theory without ever needing to claim that social facts are ‘joints in nature’. This allows socially constructed phenomena to count as real without requiring mind-independence. Banks and borders, marriages and presidential elections, for instance, are not real because they are mind-independent regularities akin to gravity, they are real because statements about them are true in virtue of collective practices, and those truths have actual causal and normative consequences. A truthmaker approach to social ontology makes this claim via the argument that objective social facts are made true by these socially constructed parts of reality (Griffith 2025). Treating social facts as real only if they are mind-independent amounts to a

commitment to a forced anti-realism that ignores the robust quality of the social world.

3 Mind-Independence

The tension between social constructionism and realism depends on an affirmative answer to (Q3): ‘Is a statement true in virtue of mind-independent facts?’ The affirmative answer here aligns with a traditional, strong version of metaphysical realism that views mind-independence as the mark of reality. Under such a view, anything that is socially constructed, i.e. anything dependent on human attitudes, activity, or social interaction, is regarded as not real.

The appeal to mind-independence could be motivated by epistemic concerns. Then the thought would be that only mind-independent facts can provide an objective constraint on belief, and if social facts are mind-dependent, they lack the kind of objectivity required for knowledge. However, this misses the specific epistemology of human kinds. Rather than being a defect, the dependence of many institutional facts on collective beliefs and intentions can provide agents with a distinctive epistemic advantage regarding the relevant institutions, reducing, though not eliminating, the scope for radical ignorance and error of the kind that arises in the natural sciences (see Thomasson 2003). This traditional inference conflates two distinct notions of dependence. Epistemic objectivity does not require metaphysical independence from all mental activity. It requires only that the truth of a claim not be reducible to the attitudes of any particular subject. Once this distinction is in place, it becomes possible to see how social facts can be both mind-dependent and non-subjective. Their dependence is mediated by shared practices that are themselves subject to norms against which beliefs and assertions can be assessed. In this way, social facts can constrain thought and action without being mind-independent in the traditional sense.

However, we should view constructionism not as in opposition to realism, but as a specific account of the ontological status of social facts. To maintain a plausible realism in the social domain, we must jettison the requirement of mind-independence. In this vein, Rebecca Mason (2021)

argues that social kinds are *essentially* mind-dependent, i.e. it is part of their very nature that their existence depends, at least partially, on certain mental states. Yet, acknowledging this essential dependence does not force us into anti-realism. Social facts are clearly mind-dependent in their origin and maintenance, but more accurately they should be regarded as *practice-dependent*. For instance, it is clear that the fact that Ursula von der Leyen is President of the European Commission is mind-dependent, as it depends on manifold and complex structures of legal and political arrangements which constitute and maintain it. That supports treating social facts as truth-apt (Q1) and, in some cases, true (Q2). By answering in the negative to the requirement of mind-independence (Q3), we do not abandon realism altogether.

According to the resulting variant of realism, social facts are constructed through social practices, but they nonetheless constitute robust parts of the reality we encounter. They shape social expectations and provide normative guidance not subject to individual fiat. In this sense, constructionism should be cast not as a denial of robust existence, but as a description of how social reality is built and sustained. Once constructionism is understood as a thesis about practice-dependence, it is no longer in opposition to minimal realism.

On the broad account of practice-dependence outlined here, social facts such as a person's status as president, the closing time of a bank, or the value of a currency obtain in virtue of ongoing social practices. Their dependence on these practices is constitutive rather than merely causal. These practices involve collective conceptualizations and specific ways of interacting that ground social facts. Crucially, this dependence does not make the facts subjective in a way that would allow an individual to simply think them away. While they are mind-dependent in the sense that they require human minds to exist, they are robustly out there for any individual participant within that society.

To understand this form of mind-dependence, we must move beyond a simple binary. In the natural sciences, such as physics and chemistry, the ontological status of an entity might indeed best be cast in terms of its mind-independence, a standard that some philosophers of science continue to defend for natural kinds (see Brzović 2024). But in the social world, if there were no agents engaged in interaction or forming shared attitudes, there would be no banks, clubs, or elections. This does not make them less

real, it locates their reality differently. As Elizabeth Barnes notes, social structures are ‘as real as anything’, even though they are joints in the social world rather than joints in nature (Barnes 2017). Our collective social activity is responsible for their persistence, but they are no less real as a result.

We can also refine our understanding of mind-dependence by contrasting causal and constitutive dependence. For example, a social fact might be causally dependent on past actions, yet remain independent of current observers, and this marks a crucial difference between subjectivist and intersubjectivist accounts. While a subjectivist might see a social fact as true in virtue of an individual’s attitudes, an intersubjectivist sees it as true in virtue of a social practice or attitudes shared within a community. According to the intersubjectivist view, social kinds and institutional facts are quasi-objective, in the sense that they do not go away just because one individual stops thinking about them or changes their mind.

The notion of intersubjectivity can be further elucidated by considering the role of coordination. Many social facts arise from the need to coordinate behavior across agents. Such coordination problems admit of multiple solutions, but once a solution is established, it acquires a degree of stability that is not merely psychological but normative. Deviations from established practices are liable to criticism, sanction, or exclusion. The resulting structure is not reducible to individual intentions, even though it depends on them in a general sense.

The same point explains the resilience of social facts. Even when individuals reject a given norm, they may find themselves constrained by the continued adherence of others, which gives social norms and facts a form of resilience that can be taken as evidence of their reality. Such resilience or inertia is not a mark of mind-independence, but of the robustness of the underlying practices.

Moreover, social reality contains heterogeneous elements with varying degrees of mind-dependence. Khalidi (2015) here instructively distinguishes between social kinds whose type and tokens are mind-dependent, and those social kinds for which only the type is mind-dependent. For example, one can only be a president if they are collectively regarded as such – the status is mind-dependent at both the level of the category and the individual instance. However, a specific dollar bill might

exist as a token of its kind even if it is lost and no one is currently aware of it. In this latter case, the individual object retains its social status through the general existence of the social kind, even in the absence of targeted mental awareness.

Another distinction lies between emergent patterns of interaction and conferred statuses. In her developed account of social construction, Ásta (2018) defines these conferred statuses as properties that are granted, whether formally or informally, by individuals or institutions with social standing, based on perceived base traits. (Though, as Adam Hochman (2024) recently argued, the exact theoretical boundaries and relative utility between strict conferralist models and broader 'social kind' approaches constitute a site of substantive debate.) Patterns of interaction can emerge without anyone intending or even noticing them, and are therefore less directly dependent on particular attitudes than conferred statuses, which depend on acts of conferral. Both depend on human action and attitudes in a general sense, but conferral involves a further form of attitude-dependence. These distinctions support a more fine-grained taxonomy of realist views in social ontology. (Mapping this more fine-grained terrain, however, is well beyond the scope of this chapter.) Decoupling realism from mind-independence must not collapse into relativism. Some recent attempts to modify conditions for realism in social ontology, such as tying it to 'relative frame manipulability' (Tanzil 2025), risk making the reality of social facts problematically relative to specific cultural contexts, implying that a social kind is 'less real' simply because its rules are easier to change.

A definition of realism that insists on mind-independence is thus ill-suited for social ontology, but so is one that collapses into relativism. If we define realism instead through success theory (R2), we find a fully commendable definition for the social domain. This position acknowledges that social entities can have real existence and that there can be facts of the matter concerning them, even if those facts are grounded in our practices.

When an agent asserts that a social fact is real, they are pointing to its non-redundancy and its causal and normative efficacy. Causal efficacy is often regarded as a mark of being real, i.e. if something makes a causal difference in the world, we have reason to think it is real. Normativity works similarly in this regard, for if a social arrangement generates oughts and obligations, it has a claim to reality. For example, many national

borders exert causal efficacy by physically halting the movement of persons and goods, while the social institution of a contract demonstrates normative efficacy by generating legally and socially binding obligations to fulfill its terms. Or consider how the fact of racial categorization causally impacts life expectancy and wealth distribution, while the conferred status of citizenship generates an extensive set of civic duties and exclusive rights. These social facts make a difference not because they are made of mind-independent matter, but because they are non-redundant in our explanations of the social world.

Social construction describes the dependence relations in virtue of which such facts obtain. Adopting this realism does not imply a static commitment to a completed construction but an engagement with the social arrangements we constantly maintain through collective activity. If social reality is thus practice-dependent, questions of ontology and social change are closely connected. If our practices ground these facts, we are also in a position to transform them.

Once it is acknowledged that social facts (and social kinds) are practice-dependent constructs and as such real in the specified sense, we can approach social critique in a specific way. We no longer need to argue that social facts are ‘unreal’ or ‘mere illusions’ in order to justify changing them. On the contrary, we must insist on their reality in order to explain why they exert such pressure on our lives and why their transformation requires a change in practice. For example, to treat social oppression as unreal simply because it is a construct would be to lose the ability to analyze the very mechanisms we seek to dismantle. A success-theory realism allows us to assert that an oppressive structure is real, and precisely because its reality is grounded in our practices, we have the power to transform it.

Decoupling realism from mind-independence allows for a descriptive ontology of current social structures without precluding their normative critique and subsequent change. It supplements accounts centered on fixed, mind-independent entities with an account of practice-dependent entities. This conception permits critique of existing practices without denying (or diminishing) their reality.

4 Social Change

We can now return to another aspect of the parent's rhetorical maneuver in the McPhail cartoon. By appealing to the reality of bedtime, as opposed to the constructedness of the social kinds they mention, the parent treats a social arrangement as non-negotiable. In this view, reality is treated as synonymous with immutability. However, the parent obscures a normative demand by presenting a revisable practice as immutable. If bedtime is a social fact, an arrangement of time, authority, and domestic practice, it is just as much a construct as the other items on the parent's list. The parent's insistence on its reality is not a hint at something mind-independent, but a strategy to shield a specific social practice from the child's demand for change.

To understand how social facts can be both real and mutable, it may be useful to distinguish between different kinds of change. Some changes are merely superficial, affecting the expression of a practice without altering its underlying structure. Others are more fundamental, involving a reconfiguration of the norms and expectations that constitute the practice. A theory of social change must be able to account for both kinds of transformation, and it should reject the inference from reality to immutability. I would contend that social facts are real precisely in the ways discussed above – statements about them are truth-apt, some of them are indeed true, and the facts they denote exert causal and normative force – but this reality does not exempt them from change. In fact, they are mutable exactly due to the specific way in which they are real, namely through their practice-dependence. We must emphasize that social reality is non-fundamental and grounded in human attitudes, actions, and practices (for discussion see Laitinen/Schweikard 2025). There is good reason to reject the 'fundamentalist' conception of metaphysics that treats only the absolute bottom layer of reality as worthy of study. The non-fundamental, constructed social world is an equally robust and legitimate domain of metaphysical inquiry (Mason/Ritchie 2020). This makes room for a realist account of social facts without appeal to mind-independence. We can adopt a cognitivist and success-theoretical stance towards statements about social facts, but it is the negative answer to the question concerning mind-independence that provides the key to understanding social change.

By untying the real from the mind-independent, we recognize social structures as joints in the social world (see again Barnes 2017). The

persistence of such structures depends on collective social activity. Social practices are not static: they are constantly performed, reinforced, and potentially modified, and changes in these practices can affect what kinds of social entities exist (see Root 2000). This means that the maintenance of a social institution such as money, or of a social kind such as gender, is a continuous process rather than a fixed state. In many cases, individuals are aware of how they are classified and react to these categorisations by altering their behaviour or self-conception. This looping effect can force the categories themselves to adapt, making some social kinds 'moving targets' (Hacking 2007). If we accept a minimal realism (R2) as characterized above, we admit that there are truths about our social world. For instance, it is a fact that certain groups face racial discrimination or that a specific currency holds value. These are not mere figments of imagination, they have causal and normative consequences. Yet, because these facts are true in virtue of mind-dependent practices, they are in principle mutable.

A central motivation for social constructionism is to show that phenomena caused or sustained by social factors could have been, or might still become, otherwise (Mallon 2016). This does not mean that such phenomena are straightforwardly under collective control. A distinction between 'opaque' social change and 'intentional' social transformation therefore becomes crucial. Whereas many social structures undergo transformations 'behind our backs' through the aggregate drift of individual actions, a realist account of social change can also make room for their intentional transformation through collective agency. Because these facts are maintained through collective activity, agents may, under suitable conditions, revise the practices that ground them. The proposed realism helps to explain how such agency is possible. Effective intentional transformation requires an adequate account of the structure and of the causal mechanisms through which it is maintained. If we treated racial oppression as 'unreal' simply because it is a construct, we would lose the ability to analyze the very mechanisms we seek to dismantle. Acknowledging this reality is important for emancipatory politics because maintaining the objectivity of social truth is not a form of oppression but a condition of liberation (cf. Griffith 2025).

Intentional social transformation targets sustained social practices from within the practice. As Haslanger (ms.) argues in her work on structural

disruption, these practices are the crucial nodes where cultural schemas (our shared concepts and norms) intersect with material resources. Therefore, genuine transformation cannot rely on changing minds alone, it requires a simultaneous reconfiguration of both our ideological schemas and the actual distribution of resources. Such efforts typically require questioning the status quo, raising the requisite collective awareness, the articulation of new norms, the mobilization of support, and the negotiation of resistance. The success of such efforts depends on a range of factors, including institutional structures and power relations. But a realist account of social facts provides the conceptual resources needed to analyze these processes. By treating social structures as real in the sense of practice-dependence, we can investigate the mechanisms through which they are maintained and transformed. This includes examining the role of recognition, the distribution of authority, and the conditions under which practices can be successfully revised.

The account of social change sketched above has a normative dimension. If social facts are mutable because they depend on collective practices, questions arise about responsibility for the persistence and transformation of unjust arrangements. Such responsibility need not be shared equally by all participants in a practice. It may rest more heavily with those who actively perpetrate injustice, who benefit from it, or who are best placed to change the relevant practice. How we individuate the (collective) agents of change, and whether these will be formal, authoritative bodies or not, will depend on whether we focus on a practice within a certain community and whether we focus on a particular constellation or a more general phenomenon. My point here is that acknowledging the nature and reality of social structures is a prerequisite to effective social critique and transformation (see Haslanger 2012).

The child in McPhail's cartoon can be understood as suspecting that the parent invokes 'reality' to block negotiation. The position I defend here rejects the forced choice implied by the parent's argument. We can treat social structures as real without treating them as immutable, and acknowledge their constructed character without treating them as mere illusions. Instead, we adopt a realism that acknowledges the nature of the social world while insisting on its mutability. In this framework, the very

features that the parent in the cartoon uses to suggest that gender or money are constructs are the features that make social change possible.

Realist constructionism argues that social ontology concerns practices that are ongoing and revisable rather than fixed entities (Hacking 2007). In this way, it clarifies that social change is achieved by revising the practices that constitute social reality, not by overcoming reality itself. The reality of social facts is on this account a precondition for meaningful critique and transformation as opposed to an obstacle to it.

5 Conclusion

The argument of this chapter has been that the perceived antagonism between realism and constructionism rests upon a failure to distinguish between different orders of metaphysical commitment. As it turns out, the parent's error in McPhail's cartoon lies not in asserting that bedtime is real, but in claiming that its reality is different from that of the mentioned constructs, that it is practice-independent, at least in the sense that it is beyond the negotiation the child seeks. However, as soon as we adopt a more differentiated account of realism, we find that a minimal realism (R2) is perfectly compatible with the view that social facts are generally practice-dependent.

This shift is consequential for our understanding of social change. It indicates that the reality of social facts need not be understood as immutability, and that their constructed character does not make them illusory. Social facts can be both real and practice-dependent. Because these facts are grounded in ongoing human interactions and collective conceptualizations, they possess a causal and normative significance. This dual character of being both practice-dependent and effectively real is not a defect, but a defining feature of the social domain. Yet the very same ground that grants them reality also ensures their mutability. If a present social fact obtains in virtue of an ongoing practice, it will cease to obtain if that practice is sufficiently transformed. Such facts are therefore in principle mutable, even when changing them is difficult or beyond the power of any particular agent.

We can therefore side with the child in the cartoon and interpret their reluctance as the recognition of a practice-dependent reality that is currently being enforced by parental authority. Bedtime is as real *and* practice-dependent as any social fact and arrangement, and as such it is not beyond the scope of transformation. In the social domain, being real does not mean being fixed by nature. Rather, it means that there is a fact of the matter, grounded in shared practices, that constrains our actions and explanations. We may not have reason to be optimistic about *this* negotiation, but that has more to do with the pragmatics of child-rearing than with the nature of social facts.

We can then make explicit the implications this view has for both methodology and critique. Debates about the reality of social kinds should not be framed in terms of their independence from human practices, but in terms of their role within those practices. Questions about existence, grounding, and efficacy must be addressed in a manner sensitive to the distinctive features of the social domain. Social facts are thus neither illusory nor immutable. As Tanzil (2025) has recently argued regarding socially constructed properties, by aligning the status of social facts with the broader metaphysics of universals or tropes, we can draw on traditional arguments for existence without constantly having to revise the foundational conditions for realism itself. They are joints in the social world rather than joints in nature, real in virtue of their role in structuring action, yet essentially practice-dependent.

Finally, this framework clarifies the relation between ontology and critique. To criticize a social structure is not to deny its reality, but to challenge the practices that constitute it. Following Haslanger's framework, effective critique must target both the conceptual and material scaffolding that holds the structure in place (see Haslanger, ms.). A realist understanding of the structure strengthens such criticism by identifying what sustains the structure and how it exerts causal and normative effects. The realism required for social ontology must accommodate the constructed nature of its objects. By identifying construction as a mode of social dependence rather than its negation, we provide a framework that accounts for both the stability and mutability or revisability of social facts. Whether or not we choose to alter those practices is a separate, normative question, but the present ontology establishes that such alteration is in principle

possible, although whether it can be achieved in practice is a further question.

References

Ásta. 2018. *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*. Oxford: Oxford University Press.⏎

Bach, Theodore. 2023. ‘The Importance of Realism about Gender Kinds: Lessons from Beauvoir’, *Analyse & Kritik*, 45(2): 269–295.

Barnes, Elizabeth. 2017. ‘Realism and Social Structure’, *Philosophical Studies*, 174(10): 2417–2433.⏎

Brzović, Zdenka. 2024. ‘The Many Faces of Realism about Natural Kinds’, *Journal for General Philosophy of Science*, 55: 289–307.⏎

Epstein, Brian. 2018. ‘Social Ontology’, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition). https://plato.stanford.edu/archives/sum2018/entries/social-ontology/.⏎

Griffith, Aaron. 2025. *Truth and Social Reality: A Metaphysical Inquiry*. New York: Oxford University Press.⏎

Hacking, Ian. 2007. ‘Kinds of People: Moving Targets’, *Proceedings of the British Academy*, 151: 285–318.⏎

Haslanger, Sally. 2005. ‘What Are We Talking About? The Semantics and Politics of Social Kinds’, *Hypatia*, 20(4): 10–26.⏎

Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.⏎
Haslanger, Sally. (ms.). ‘How to Change a Social Structure’. Available at https://www.ucl.ac.uk/laws/sites/laws/files/haslanger_how_to_change_a_social_structure_ucl.pdf.⏎
Hochman, Adam. 2024. ‘Social Kind vs. Conferralist Approaches to Social Ontology: “Race” as a Case Study’, *Inquiry*. https://doi.org/10.1080/0020174X.2024.2406883.⏎
Khalidi, Mohammad Ali. 2015. *Natural Categories and Human Kinds*. Cambridge: Cambridge University Press.⏎
Laitinen, Arto and David P. Schweikard. 2025. ‘Defining Realism in Social Ontology’, in A. Laitinen, M. Keinänen, J. Reinikainen and A. Honkasalo (eds), *Language, Truth, and Reality: Philosophical Essays in Honour of Panu Raatikainen*, pp. 196–226. Tampere: Tampere University Press.⏎
Mäki, Uskali. 2008. ‘Scientific Realism and Ontology’, in St. N. Durlauf and L.E. Blume (eds), *The New Palgrave Dictionary of Economics*, pp. 334–341. Basingstoke: Palgrave Macmillan.⏎
Mallon, Ron. 2016. *The Construction of Human Kinds*. New York: Oxford University Press.⏎
Mason, Rebecca. 2021. ‘Social Kinds Are Essentially Mind-Dependent’, *Philosophical Studies*, 178(12): 3975–3994.⏎
Mason, Rebecca and Katherine Ritchie. 2020. ‘Social Ontology’, in R. Bliss and J. T. M. Miller (eds), *The Routledge*

Handbook of Metametaphysics, pp. 312–324. London: Routledge.⏎
McPhail, Will. 2022. *Love & Vermin: A Collection of Cartoons*. New York: Mariner Books.
Mills, Charles W. 1998. *Blackness Visible*. Ithaca: Cornell University Press.⏎
Root, Michael. 2000. ‘How We Divide the World’, *Philosophy of Science*, 67 (Supplement): S628–S639.⏎
Sayre-McCord, Geoffrey. 1986. ‘The Many Moral Realisms’, *The Southern Journal of Philosophy*, XXIV: 1–22.⏎
Schmid, Hans Bernhard. 2023. *We, Together: The Social Ontology of Us*. New York: Oxford University Press.
Tanzil, Vincent. 2025. ‘Realism and Antirealism About Socially Constructed Properties’, *Philosophy Compass*, 20: e70060.⏎

OceanofPDF.com

2 Institutions and Normativity

Needs, Joint Action, and Reciprocity

Seumas Miller

DOI: 10.4324/9781003612827-3

1 Introduction

Social institutions have a normative dimension or, perhaps, multiple normative dimensions. Consider the following, presumably true, general claims. Other things being equal, the occupants of institutional roles *ought* to perform those actions that are the means to realise their institutional ends. Here what might be termed 'the instrumental ought' is in play. Moreover, in pursuing these ends they *ought* to respect the moral rights of others. Here what might be termed 'the moral ought' is in play. Normativity in its widest sense pertains to what ought to be the case, what ought to be done, what benefits one ought to receive and so on. However, there is the instrumental 'ought' and the moral 'ought'. But there is also the prudential 'ought', the aesthetic 'ought', and so on. The concern in this chapter is with social institutions and normativity but, more specifically, with the source, or rather sources, of moral normativity in social institutions. There are at least three such sources, or so I will argue: needs, joint productive activity, and reciprocity.

In addressing the issue of the sources of moral normativity in social institutions, I assume social institutions are organisations (and structures of organisations). I also assume a certain account of social institutions, namely, the normative teleological account of social institutions (Miller 2010). Social institutions in the sense of organisations should be distinguished from conventions, norms, regulations, and laws, i.e., social rules (and structured sets thereof) that often govern individual (as opposed

to joint) action. Consider the rules of the road, e.g., driving on the left-hand side of the road, giving way to cars on one's right side at intersections and driving at or under the speed limit. Each single driver in so far as she is simply driving her car in a given direction is engaging in an individual action rather than a joint action; she is not, for example, sharing the car's control with a second driver in her car. However, this essentially individual action is constrained by the rules in various ways, e.g. she drives on the left-hand side and at or under the speed limit. Such rule-governed *individual* action is not necessarily rule-governed *joint* action. In driving from A to B, x is intending to get to B while y in driving from B to A on the same road is intending to get to A. Here x and y do not have a common destination, and x can happily drive to A irrespective of whether y drives to B. However, it may well be the case that compliance with the rules is joint compliance, i.e., each conforms interdependently with conformity by the others, e.g., each motorist drives on the left if the others drive on the left (and would drive on the right if others did so). Moreover, at any point at which two cars are passing one another going in opposite directions there is, arguably, a joint action since each driver, A, seeks to avoid a collision with the other driver, B, by ensuring that A is on the opposite side of the road to B (and vice-versa for B), i.e., A and B coordinate to achieve the collective end of avoiding a collision (Miller 1992; Miller 2001).

As with the rules of the road, the criminal law consists of a structured set of rules (based, in large part, on prior moral principles, e.g., don't kill innocent persons) governing individual actions (and, in some cases, discrete joint actions, e.g., two persons jointly killing a third person), albeit in the case of the criminal law many of the individual actions in question are prohibited (as opposed to the manner of their performance being prohibited). However, in the case of the criminal law (and the rules of the road in so far as they are laws) we can distinguish between primary and secondary laws; secondary laws being laws of procedure governing the primary laws, e.g., secondary rules determining what counts as a valid (primary) law (Hart 1961).

By contrast with rules and structured sets of rules governing individual actions (including legal systems comprised of primary and secondary rules), organisations at their core consist of joint activity, i.e., actions performed to achieve a shared end (collective end in my parlance). Moreover, these joint

actions are themselves components of larger-scale joint actions, i.e., multi-layered structures of joint action (see next section). The raison d'etre or institutional purpose of a car manufacturing company is to design and produce cars and the activities of the managers and employees are structured to achieve this purpose or collective end. Thus, the factory worker who fits the wheels onto the axles would not be doing so if other workers were not mounting the chassis onto the frame. Moreover, the joint actions of the workers on the car assembly line combine with the joint actions of the designers and makers of the component parts in the service of the collective end of the production of drivable cars.

On this account of social institutions, they consist in large part in *multi-layered joint activity* in the service of collective ends which are also collective goods (Miller 1992; Miller 2001: Ch. 5; Miller 2010: Chs. 1 & 2). Thus, a team of detectives engages in the joint activity of investigating a crime in order to realise the proximate collective end of identifying and arresting an offender and do so—in conjunction with the investigative and other activities of other members of the police organisation—in order to realise the ultimate collective end of maintaining law and order which is a collective good.

Moreover, I adopt a realist account of normativity in general (Wedgwood 2007) and a moral realist account, in particular. On this account, moral claims, such as that one ought to respect the moral rights of others, are truth claims (and, therefore, have truth conditions). Moreover, if sincerely made, such claims express moral beliefs and judgements, i.e., cognitive mental states. Further, moral beliefs are motivational states (Benn 1988) in the sense that they motivate the corresponding intentions and intentional actions (Wedgwood 2007). Thus, if someone believes that they ought to keep their promise to pay their debt today, then (other things being equal, e.g. they are not suffering from akrasia), they will form an intention to pay their debt today and pay it (Wedgwood 2007: 34).

In Section 2, I elaborate my favoured normative teleological account of social institutions. In Sections 3, 4 and 5, I discuss needs, joint productive activity and reciprocity, respectively, as sources of moral normativity in social institutions.

2 Social Institutions

As mentioned above, on my favoured normative teleological account (Miller 2010), social institutions consist in large part in multi-layered structures of joint activity directed to the realisation of a collective end(s) which is a collective good.

2.1 Joint Action

The central concept in the normative teleological account of social institutions is that of *joint action* (Miller 1992; Miller 2001: Ch. 2; Miller 2007; Miller 2010: 41–46). Joint actions are actions involving a number of agents performing interdependent actions in order to realise some common goal (collective end). Examples of joint action are: two people dancing together, a number of tradesmen building a house, and a group of robbers burgling a house. Joint action is to be distinguished from individual action on the one hand, and from the 'actions' of corporate bodies on the other. Thus, an individual walking down the road, driving a car or shooting at a target are instances of individual action. A nation declaring war or a government taking legal action against a public company are instances of *corporate* action.

On the collective end theory (CET) of joint action (Miller 1992; Miller 2001: Ch. 2; Miller 2007; Miller 2010: 41–46), basic joint action consists of: (1) a number of singular actions; (2) relations between these singular actions, e.g. of interdependence; (3) a collective end at which these singular actions are directed. Moreover, the constitutive attitudes involved in joint actions, e.g., intentions, beliefs and ends, are individual attitudes; there are no sui generis collective intentions (sometimes referred to as we-intentions) (Searle 1990; Gilbert 1992).

CET is an individualist account of joint action that not only eschews sui generis we-attitudes but also the ascription of mental states to collective entities per se. In doing so it does not need to hold the implausible view that there are no irreducibly collective entities, e.g., Australia, the United States, Finland and the European Union. Clearly there are irreducibly collective entities, and such entities can and do figure in the content of individual attitudes, e.g., John Searle believes that Finland, but not Australia or the

United States, is a member of the European Union. However, the existence of collective entities does not entail the existence of collective *agents*, and specifically, it does not entail the existence of supra-individual mental attitudes.

A related error often made among those who attack individualist accounts of joint action is to confuse *atomistic* individualism with *relational* individualism. Collective ends are *shared interdependent* ends that a single individual could not realise acting alone or, at least, could not do so except with great difficulty.

CET is the theory that joint actions are actions directed to the realisation of a collective end. However, this notion of a collective end is a construction out of the prior notion of an individual end. Roughly speaking, a collective end is an individual end more than one agent has, and which is such that, if it is realised, it is realised by all, or most, of the actions of the participating agents; the individual action of any given agent is only part of the means by which the end is realised. So, a joint action simply consists of at least two individual actions directed to the realisation of a collective end.

Moreover, the notion of a collective end is an individualist notion. The realisation of the collective end is the bringing into existence of a state of affairs. Each agent has this state of affairs as an individual end. (It is also a state of affairs aimed at under more or less the same description by each agent.) So, a collective end is a species of individual end. The individual ends involved in a joint action are realised by *one single* state of affairs; so, the coming into existence of the one single state of affairs in question constitutes the realisation of *all* of the individual ends that exist in the heads of each of the participating agents. For another thing, although collective ends are shared (individual) ends, they are ends that are interdependent.

Three further points about collective ends are as follows. Collective ends can be unconsciously pursued; and they have not necessarily been at any time explicitly formulated in the minds of those pursuing them. Collective ends can be *implicit* in the behaviour and attitudes of agents without ceasing to be ends as such. Further, in the case of a collective end pursued over a long period of time, e.g., by members of an institution over generations, the collective end can be *latent* at a specific point in time, i.e., it is not actually being pursued, explicitly or implicitly, at that point in time. However, it

does not thereby cease to be an end of that institution—which is to say, of those persons—even at those times when it is not being actively pursued.

Thirdly, a collective end is not an intention. On CET ends are distinguished from intentions. The notion of an end is coordinate with that of a means, but this is not so for intentions. To say that an action was intended is not in itself to say that the action is a means or that it is an end. Indeed, some intended actions do not involve means or ends.

A final point about joint actions pertains to normativity. It can be argued that joint actions are inherently normative phenomena and that many individualists deny this. Here we need to distinguish types of normativity. On any account, individualist or otherwise, joint actions involve beliefs, intentions, ends and the like, i.e., attitudes that are normative in the weak sense that they are subject to some form of evaluation, e.g., truth and falsity or success and failure. This much, individualists can and do acknowledge. Moreover, on any account, individualist or otherwise, many joint actions are normative in some stronger sense, e.g., possess institutional or moral properties. However, the critical question is whether all joint actions per se are *necessarily* normative in some stronger sense that might be antithetical to individualism or, at least, to the form of individualism being proposed here, e.g., *necessarily* have institutional or moral properties by virtue of being actions or by virtue of being joint actions, in particular. Individualists of the sort favoured here deny that joint actions necessarily have institutional or moral properties and have offered counterarguments and counterexamples by way of support (Miller 2001: 85–90; Miller and Makela 2005; Miller 2007).

2.2 Multilayered Structures of Joint Action

Organisational action typically consists in multi-layered structures of joint actions (Miller 2001; Miller 2024). One illustration of the notion of a multi-layered structure of joint actions is an armed force composed of different units fighting a battle. Suppose that a number of joint actions are severally necessary and jointly sufficient to achieve some further collective end not achievable by any one of these joint actions and that each of these joint actions is performed by a different unit of the armed force: an infantry unit, a flight of planes, and a mortar squad. Thus, firstly, there is the joint action

of the members of the mortar squad which consists of destroying enemy gun emplacements that are attacking the infantry platoon; secondly, there is the joint action consisting of several military planes jointly providing air cover for the mortar squad and the infantry platoon; and, thirdly, there is the joint action of the members of the infantry platoon taking and holding the ground. These three joint actions are, let us assume, severally necessary and jointly sufficient to achieve the ultimate collective end of defeating the enemy; as such, these three 'actions' (each of which is a joint action) constitute an additional single (complex) joint action, namely, the joint action of defeating the enemy. Accordingly, there is a multi-layered structure of joint action comprised of three joint actions that are elements of a complex joint action, i.e. a fourth joint action comprised of the three elements.

According to the normative-teleological theory (Miller 2010), institutions are essentially joint enterprises (engaged in multi-layered structures of joint action) in the service of collective ends which are, normatively speaking, collective goods. Thus, universities are joint enterprises in the service of the collective good of knowledge, while the housing industry is a joint enterprise in the service of providing an adequate sustainable supply of housing of reasonable quality and at a reasonable price (Miller 2010: Ch. 10).

Clearly, different institutions exist to provide different collective goods. So, there is a need to develop *special* normative theories of different social institutions (as opposed to the *general* normative theory of social institutions), such as a normative theory of universities, a normative theory of the housing industry, a normative theory of policing, a normative theory of banks, and so on.

However, there is a further point to be made. According to the normative theory of social institutions, our most important social institutions have, or ought to have, as their collective ends the production or maintenance of collective goods which frequently consist in the satisfaction of aggregate needs, such as food, shelter, healthcare, security and education.

Importantly, *aggregate* needs (collective goods) in a community generate collective—understood as *joint* (Miller 2001: Ch. 8; Miller 2006; Miller 2010)—moral responsibilities on the part of members of the community to provide for those needs. Moreover, the principal means, at

least in modern societies, to discharge these joint moral responsibilities to provide food, security, education etc. are institutions. Further, the institutional rights and duties of members of these institutions, such as agribusinesses, police services, universities etc., are derived (in large part) from the collective goods (aggregate needs) that these institutions exist, or ought to exist, to provide. Note that since these *institutional* rights and duties are derived from morally required (collective) goods, they are typically also *moral* rights and duties.

To reiterate: according to the normative-teleological theory, institutions are essentially joint enterprises that produce or maintain collective goods, and many of these collective goods are aggregate needs. Moreover, these joint enterprises typically consist of multi-layered structures of joint action.

3 Needs

One source of the normativity of institutions (thus understood) is the collective good(s) that is the raison d'etre of an institution and, in particular, collective goods that, as we have just seen, consist in the satisfaction of aggregate needs; the needs of members of populations for food, housing, public health, security, and so on.

Importantly, needs stand in contrast with desires or preferences (Wiggins 1987). Firstly, unlike desires, needs are objective phenomena. Thus, food is an objectively necessary condition for the continuation of human life irrespective of whether it is subjectively desired or not.

Secondly, again unlike desires, needs are limited. Thus, a limited quantum of food is sufficient for one's needs, whereas one's desire for wealth (or, indeed, for food in the case of a glutton) is not limited (or, at least, not necessarily limited by one's desire qua desire, even if limited by, for instance, one's capacity).

Thirdly, and most importantly for our purposes here, needs (unlike desires) are generative of moral obligations. This is because one is harmed if one's needs are not met. For instance, if one is deprived of food or water, one will become physically frail and eventually die. Accordingly, if I have a need which I cannot satisfy myself, then I have a legitimate moral claim against those who can satisfy it. That is, other things being equal, those who

can satisfy my need have a moral obligation to do so. Naturally, other things might not be equal. For instance, it might be extremely morally costly for those who can satisfy my need to do so; too costly to morally justify their doing so. Whether or not this is the case will depend on the harm consequent upon my need not being met, and on the extent of the cost to be borne by those who can satisfy it.

Moreover, if a single individual cannot satisfy my need, but joint action on the part of multiple individuals can satisfy it, then these individuals may have a joint moral responsibility to satisfy it. This is especially the case if there are large numbers of individuals with the same need as mine and this aggregate need can be satisfied by the joint action of others. Indeed, in such cases, supposing the aggregate need in question is ongoing, there may well be a joint moral responsibility to establish an institution whose occupants jointly act to satisfy the ongoing aggregate need in question.

In this connection, consider, for instance, a fire crew. Members of a fire crew have a positive moral duty to put out an incipient grass fire in the bush which—given the very high temperature and strong winds—will, if not extinguished, develop into a full-scale bush fire that will destroy property, and perhaps even lead to loss of life. Here it is not simply that the members of the fire crew, and indeed anyone else, would be doing a good thing if they put out the grass fire. Indeed, unlike ordinary members of the public, the members of the fire crew are morally *required* to extinguish the fire; it is their morally required institutional duty to extinguish the fire. So now the question to be asked is: What is the moral basis for this institutional duty?

Presumably, there is a prior moral requirement to see to it that, for example, there are fire brigades and firemen to extinguish fires in order to avert the very serious harms consequent upon uncontained fires. However, in addition to the duty of an individual to assist where he or she is the only one able to do so and there is minimal cost involved, there are also duties or responsibilities of the members of groups to so assist; in short, there are joint moral responsibilities.

So, the need of members of groups for various forms of assistance that can only be adequately rendered by members of other groups generates joint moral responsibilities on the part of the members of the latter groups to so assist. Moreover, where such joint responsibilities can most effectively be discharged by establishing institutions and institutional roles the

constitutive *institutional duties* of which consist of providing such aid, e.g., fire brigades and firemen, then members of the group who have the prior joint moral responsibility may have a derivative joint moral responsibility to establish and support such institutions.

Moreover, these institutional duties are also moral duties. For these institutional duties derive from the larger moral purpose served by the institution and its constitutive institutional roles, e.g., the moral requirement to avert the destruction of life and property caused by fires. Accordingly, the prior joint moral responsibility to assist devolves to the relevant institutional actors; it is now their joint moral and institutional duty to (in our example) fight fires.

Thus, we can readily see that the collective goods that are the raison d'etre for social institutions and that consist of aggregates of needs, straightforwardly generate moral rights and duties on the part of the role occupants of those institutions.

Here those occupational groups that are typically referred to as the professions, such as doctors and lawyers, are salient. For the collective ends, indeed collective goods, of these occupational groups are, I suggest, the satisfaction of needs (Alexandra and Miller 1996; Miller 2010: Ch. 6). This is obvious in the case of doctors whose raison d'etre or collective good is the achievement, restoration and retention of physical health. However, justice is also a need and is the raison d'etre for the profession of lawyer. Again, financial health is the raison d'etre of finance professionals such as accountants. As this latter example perhaps indicates, when we talk about needs, we are not necessarily always referring to the needs which people have in virtue of being human (such as the need for good health), but rather also to needs which they have in virtue of their participation in historically localised forms of social organisation. (There is no need for financial well-being in a society which practises primitive communism, for example.) As noted above, someone can be said to have a need when they will suffer harm if they fail to obtain the object of that need. However, we can distinguish between fundamental needs and other needs; the failure to satisfy fundamental needs will cause *significant* harm. Drawing this distinction enables us to further refine the definition of the professions and, for that matter, the definitions of many, if not all, social institutions the raison d'etre of which is the satisfaction of aggregate needs.

Not all needs will count as fundamental in this sense, since we may legitimately claim that we need something when it is the only available means to the achievement of some desired state of affairs, but where failure to achieve that state of affairs will not cause us to suffer significant harm. David Wiggins gives the example of the person who needs two hundred pounds to buy a suit—yet does not have a fundamental need to buy the suit (Wiggins 1987: esp. 7–9). Often it will be difficult to distinguish fundamental from non-fundamental needs—is university education a fundamental need, for instance? It is also true that often it is *not* difficult to classify a need as fundamental: if I am suffering from a life-threatening disease then I have a fundamental need for medical aid; and if I live in a modern industrialised society then I have a fundamental need for primary and secondary education, at least.

Even if it is the case, as I have claimed, that the satisfaction of fundamental needs is the end of professional activity, professionals are not the only workers who satisfy such needs. Eating food is a fundamental need —nevertheless, a sandwich hand is not a professional. Moreover, there are many social institutions whose constitutive roles are not professions, yet which have as their raison d'etre the satisfaction of aggregate fundamental needs, including agribusinesses (farmers) and police organisations (police officers).

4 Joint Productive Activity: 'Mixing One's Labour'

Thus far it has been argued, in effect, that needs generate obligations to satisfy these needs and moral rights on the part of the needy to the needed goods. But what of those who discharge these obligations to the needy, i.e., those who produce the needed goods? The producers of these goods also have moral rights to the needed goods. They have these rights by virtue of having produced these goods, although these producers' rights must accommodate the rights of the needy. At any rate, another source of the normativity of institutions on my favoured account is the joint activity which produces (maintains or renews) the collective goods that are the raison d'etre for institutions. In jointly producing a collective good,

participants in this productive activity, thereby, have moral rights to this good or, at least, have moral rights to some reward for producing it. Here John Locke's conception of property rights is relevant and, more specifically, his notion of mixing one's labour with, for instance, the land by ploughing it and planting a crop (Locke 1980: Ch. 5). Thus, if we suppose a farming family thereby produces a wheat crop then surely the members of the family have a joint moral right to consume or sell the crop.

However, as Kogelmann (Kogelmann 2021) notes following Nozick (Nozick 1974), not all labouring produces value. Thus, to use Nozick's example, spraying pink enamel paint on a piece of driftwood is an act of labouring, but it does not make the driftwood any more valuable (Nozick, 1974: 175). It might even make it less valuable. This argument, as Kogelmann in effect states, demonstrates (assuming it is valid) that mixing one's labour must produce value if it is to generate moral rights. However, the mixing of labour that is in question here produces goods that satisfy needs; that is, the mixing of labour in question does produce value and, specifically, produces collective goods that meet aggregate needs. Accordingly, the argument does not count against acceptance of the view proffered here, namely, that those whose labour (joint labour) produces collective goods have moral rights to that good or, at least, to some reward for producing it.

To reiterate. Sometimes the end realised in joint action is not merely a collective *end*, it is also a collective moral *good*, e.g., a collective good that satisfies an aggregate need. If so, then a joint moral right may well be generated. It is easy to see why these agents, and no other agents, would have a right to such a good; they are the ones responsible for its existence, or continued existence. In this connection consider the managers and workers in a factory that produces cars that are sold for profit. Managers and workers in the factory—but not necessarily others—have a joint right to be remunerated from the sales of the cars that they jointly produced. Managers and workers might in effect have waived this joint right by virtue of their employment contracts. But this does not affect the substantive point in relation to their prior joint moral right.

It is also clear that if one participating agent has a right to the jointly produced good, then—other things being equal—so do the others. That is,

there is interdependence of moral rights with respect to the jointly produced good.

Let us look a little closer at this joint right. Firstly, what is the basis of this right? Perhaps the basis of this joint right is simply the fact that each of the participants performed some individual action, and the result of their actions was the coming into existence of some good. This is not quite correct. One does not have a joint right in virtue of performing an action which is describable independently of the actions of other agents. For example, one does not have a right to use a house that a number of persons jointly built simply in virtue of the fact that one built one of the walls. Rather one has a right in virtue of having *contributed* to the building of the house by building the wall in question. One's wall building considered by itself would not give rise to a right in respect of the house. For there might not be a house, but only a single wall.

Secondly, what is the relationship, if any, between joint rights and collective goods? The good is a realised collective end to which the participants possess a joint right. Accordingly, even if the good considered in itself is not intrinsically a *collective* good (but is rather an aggregate good), each of the individual (jointly held) rights to it is a right to it qua *collective* good. For example, two mining companies might work together to enable the buying of land and the mining of ore on the land. Assume that the deposit is too large and too inaccessible for it to be profitable for one company to go it alone. By prior agreement, the two companies might each possess an individual right to sell 50% of the ore produced. But this agreement is something additional to the joint right to the land bought and the ore produced and indeed presupposes this joint right. Imagine, contra expectations, that the ore produced could not be parcelled out in the manner envisaged in the agreement. Perhaps there turns out to be only one large buyer who will not accept two separate suppliers. If so, each individual mining company could still claim an individual (jointly held) right to the ore produced, and therefore legitimately insist on making some different agreement in relation to selling it (or perhaps no agreement).

What of the relationship between, on the one hand, the joint moral right of the producers of a good that can satisfy an aggregate need to that good and, on the other hand, the moral rights of the needy to that good? As we have seen, both categories of moral rights are in play, albeit they have

different sources; one is based on need, the other on labour expended. However, what is important to stress here is the interdependence of these two categories of moral right. Firstly, the labour expended only generates a moral right if the product of that labour has value and the value in question is, at least in the contexts in question here, based on the fact that it satisfies a need, (more specifically an aggregate need). Secondly, the moral right to have the need satisfied depends on the ability of those producing the good which satisfies the need not only being able to do so but being able to do so without excessive moral cost to themselves. They can hardly be expected to satisfy someone else's need if there is an excessive moral cost to themselves in doing so. Accordingly, the producers of the good have a moral right to the good or, at least, a moral right to be rewarded for providing it and, indeed (other things being equal) a moral right to be rewarded by those who consume it, i.e., the needy.

In short, there is an interdependence, indeed a reciprocal relationship, between the moral rights of the needy and the moral rights of the producers of the good that satisfies the needs of the needy. Let us then turn more directly to the matter of reciprocity; the third normative dimension of social institutions of interest to us.

5 Reciprocity

Thus far we have discussed two sources of normativity in social institutions, namely, needs and 'mixing one's labour', and I have suggested that there is a third source of normativity: reciprocity. As we have just seen at the end of the last section, there is a reciprocal relationship, between the moral rights of the needy and the moral rights of those who 'mix their labour', i.e., the producers of the good that satisfies the needs of the needy. However, there is a wider division of labour in play than that between the producers and consumers of any given collective good that meets an aggregate need.

Consider a division of labour across different institutions such as, for instance, a hospital and a police service. Each institution produces, maintains, renews, or otherwise contributes to a collective good; in the case of the hospital and the police service, health and security (respectively).

Normatively speaking, each such social institution enables the satisfaction of an aggregate fundamental need (e.g., the health needs of members of a population (or sub-group thereof)). Moreover, in the case of each single institution, there is the above-mentioned intra-institutional reciprocal moral rights relationship between those who provide the collective good (the joint producers of the collective good) and those whose needs are satisfied by the collective good (the consumers, as I will refer to them). However, in a contemporary society, there are multiple social institutions, including health, security, education (e.g., schools), agribusinesses and so on, each of which produces, maintains, or renews a collective good(s). Moreover, there is a complex, integrated division of labour across this multiplicity of institutions, giving rise to *inter*-institutional reciprocal moral rights (as opposed to *intra*-institutional reciprocal moral rights).

Under this inter-institutional arrangement some individuals (X1…Xn) jointly produce collective good, G1, and others (Y1…Yn) collective good, G2 (and still others (Z1…Zn) collective good, G3, and so on for other collective goods). Moreover, I suggest that X1…Xn, and Y1…Yn have reciprocal moral rights to G2 and G1 (respectively), X1…Xn and Z1…Zn to G3 and G1 (respectively) and so. That is, X1…Xn have rights to the collective goods, G2 and G3, that they did *not* produce; as do Y1…Yn to G1 and G3, and Z1…Zn to G1 and G2. Accordingly, firstly, everyone has multiple needs which are satisfied by the various different collective goods produced by the various different institutions, i.e., everyone ‘consumes’ multiple collective goods, e.g., food, health, security etc.; however, secondly, everyone (or most) are also producers, although (typically) only producers of a single collective good (or rather contributors to a single institution which produces a collective good (or at most a small number of collective goods), e.g., those who are police officers maintain the collective good of security but not the collective good of health. In short, there are reciprocal moral rights across social institutions, i.e., inter-institutional reciprocal moral rights (as well as intra-institutional reciprocal moral rights between producers and consumers within a given institution). This structure of inter-institutional reciprocal moral rights (as opposed to intra-institutional reciprocal moral rights) is based on a division of labour (described above) according to which everyone (or most) is both a (‘mixing-labour’ based, rights-bearer) producer in one institution and a

(needs-based, rights-bearer) consumer in multiple institutions. The upshot of this arrangement is that each adult member of the society (or, at least, most such members) is a specialist producer and a generalist consumer possessed of (intra-institutional and inter-institutional) reciprocal moral rights to the collective goods produced by the society's various different social institutions.

Naturally, in the actual contemporary social world there are not only multiple collective goods, each of which is produced by a different institution (or specialised set of institutions); these collective goods are often produced inter-generationally. Accordingly, the reciprocal moral rights in question reach across an entire society and across the generations. For instance, children have a moral right to an education provided by adult teachers at time T1, and at time T2, their aged teachers might have a moral right to material support (in the form of a pension paid by taxpayers) from their now adult former students.

Notice that the fact that the fulfilled needs-based moral rights in question are *aggregated* makes it trivially true that the members of the community in general enjoy these rights. Likewise, since it is moral rights that are in question, then each and every needs-based rights-bearer ought to have available to them the good to which he or she has a right. Again, if my right to basic foodstuffs is fulfilled, then I enjoy the exercise of *my* right and only I enjoy the exercise of *my* right (even if others also enjoy the exercise of theirs). It is also true that the exercise of my needs-based rights to basic foodstuffs, shelter and so on is logically consistent with the inability of others to exercise their respective rights to these goods, e.g. if I am a successful subsistence farmer but one living in a failed state in which many are starving. Moreover, others can exercise their needs-based rights without me doing so, e.g. if the needed goods are only available in a market and I cannot afford to pay for them while others can. Nevertheless, each of us, albeit indirectly, relies on the fulfilment of the needs-based rights of others in order to enjoy adequately our own needs-based rights. The reason for this is twofold.

First, in modern societies most individuals rely on social institutions, e.g. agribusinesses, manufacturers of building materials etc., operating in competitive markets, to produce the foodstuffs and other necessities to fulfil their needs-based rights. Indeed, even if they were disposed to do so, few

modern individuals are even capable of producing sufficient food, clean water, adequate shelter, medicine etc. for themselves; few of us living in modern societies are, or could easily become, subsistence farmers.

Second, most individuals rely on business organisations operating in competitive markets to provide paid jobs that: (a) enable them to pay for the basic necessities of life, and; (b) generate taxes to fund a variety of other collective goods necessary for the production and distribution of these basic necessities, e.g. transport, communications, research and training, and other infrastructure.

So, there is a complex structure of direct and indirect interdependence (as opposed to one-way dependence) and overlap between the needy and those who fulfil their needs. For example, there is direct interdependence between agribusinesses and the paying consumers of basic foodstuffs; and there is indirect interdependence between the former and all the other organisations which pay their employees and, thereby, enable them to become paying consumers of basic foodstuffs.

This *de facto* indirect, web of economic interdependence—and associated structure of intra-institutional and inter-institutional reciprocal moral rights—between producer/consumers does not, of course, necessarily encompass *all* the members of a community; e.g. there might not be any dependence of the employed on the unemployed. Nevertheless, under conditions of full employment (or near full employment in conjunction with welfare payments to the unemployed) and sufficient production of basic necessities to meet the needs of all, a *de facto* indirect, web of economic interdependence will encompass all members of the community (even if not all are producers). However, the structure of intra-institutional and inter-institutional reciprocal moral rights might not be complete if the unemployed are not producers or do not otherwise contribute. This would presumably not extinguish their needs-based moral right to, for instance, the necessities for life.

This web of economic interdependence is, of course, not of such a kind that the meeting of the needs of a single person is a necessary or sufficient condition for the meeting of the needs of any other single person, let alone of all other persons taken in aggregate. Rather, the interdependence between individuals, between small subsets of the whole community, and between individuals and small subsets is partial and incremental. Roughly speaking,

the larger the subset, the greater the dependence on it of its members (taken individually) and of individuals and subsets outside it; and the less dependent it is on any particular subset outside it (or on any small subset of itself).

Each and every member of the community has a needs-based right to the basic necessities of life. Accordingly, if a *de facto* web of economic interdependence (but not necessarily of producer/consumers) is complete—and there are adequate production levels—then the needs-based rights of all will be fulfilled. Moreover, such a completed web of economic interdependence might parallel a deontic structure of interdependent (aggregated) needs-based rights (but not necessarily of intra-institutional and inter-institutional reciprocal producer/consumer moral rights). The (aggregated) needs-based rights in question are interdependent by virtue of being joint rights. Let me explain.

As mentioned above, a needs-based right is not *per se* a jointly held right; it follows that an aggregate of needs-based rights is not necessarily a set of jointly held rights. However, a needs-based right and, likewise, an aggregate of needs-based rights, are such only in the context of the possibility of their fulfilment (either by the rights bearers themselves, or by others); one cannot have a right to something, if it is impossible (logically or practically) for it to be provided.

The context in question, i.e. a well-functioning modern economy, is one in which aggregated needs-based rights are fulfilled (and realistically can only be fulfilled) by economic institutions characterised by a (more or less) completed web of economic interdependence among the consumers of basic necessities, i.e. the bearers of the needs-based rights in question. But in that case—given that rights exist only if it is possible for them to be fulfilled—then the needs-based rights in question are *joint* rights. One member of the community in question only has a right to basic necessities if others do, and vice-versa. The institutional arrangements in question are not such that they could provide for one person, or even a small group of persons; they are designed to provide for aggregate needs, i.e. for large groups of consumers. Since it is not possible to provide for one person (or even a small group) that person cannot have a right to the basic necessities independent of others having this right. That is, the right of any one person to the basic

necessities is a jointly held right; the needs-based rights in question are joint rights.

6 Conclusion

The concern in this chapter has been with social institutions and normativity but, more specifically, with three sources of moral normativity in social institutions: needs, joint productive activity, and reciprocity.

In addressing these sources of moral normativity in social institutions, I have assumed that social institutions are organisations (and structures of organisations). I have also assumed a certain account of social institutions, namely, the normative teleological account of social institutions. On this account of social institutions, they consist in large part in *multi-layered joint activity* in the service of collective ends which are also collective goods.

The collective goods in question are often aggregated needs, e.g., the health needs of a population. Hence the relevance of needs as one of the sources of moral normativity discussed.

The collective goods in question are the product of joint activity. Moreover, I accept the Lockean view that a person has moral rights to the product of his or her labour (assuming the product has value, e.g., satisfies a need). Accordingly, a set of individuals has joint moral rights to the product of their joint activity. Hence, such joint activity is a second source of moral normativity discussed.

The third source of moral normativity discussed is that of the intra-institutional and inter-institutional reciprocal moral rights generated by members of a society who are in most instances producer/consumers within a complex, integrated structure of social institutions and, therefore, have moral rights to rewards as producers but also (as consumers) needs-based moral rights.

References

Alexandra, A. and Miller, Seumas. 1996. ‘Needs, Moral Self-consciousness and Professional Roles’, *Professional Ethics*,

5(1–2): 43–61.⏎
Benn, Stanley. 1988. *A Theory of Freedom*. Cambridge: Cambridge University Press.⏎
Gilbert, Margaret. 1992. *On Social Facts*. Princeton N.J.: Princeton University Press.⏎
Hart, H. L. A. 1961. *The Concept of Law*. Oxford: Oxford University Press.⏎
Kogelmann, Brian. 2021. ‘Lockeans against Labour-mixing’, *Politics, Philosophy and Economics*, 20(23): 251–272.⏎
Locke, John. 1980. *Second Treatise of Government*. Indianapolis, IN: Hackett Publishing.⏎
Miller, Seumas. 1992. ‘Joint Action’, *Philosophical Papers*, 21(3): 275–297.⏎
Miller, Seumas. 2001. *Social Action: A Teleological Account*. New York: Cambridge University Press.⏎
Miller, Seumas. 2006. ‘Collective Moral Responsibility: An Individualist Account’, in Peter A. French (ed.), *Midwest Studies in Philosophy*, pp. 176–193. Malden, MA: Blackwell Publishing.⏎
Miller, Seumas. 2007. ‘Joint Action: The Individual Strikes Back’, in S. L. Tsohatzidis (ed.), *Intentional Acts and Institutional Facts: Essays on John Searle’s Social Ontology*, pp. 73–92. Dordrecht: Springer Press.⏎
Miller, Seumas. 2010. *The Moral Foundations of Social Institutions: A Study in Applied Philosophy*. New York: Cambridge University Press.⏎

Miller, Seumas and Bossomaier, Terry. 2024. *Cybersecurity, Ethics and Collective Responsibility*. New York: Oxford University Press.⏎

Miller, Seumas and Makela, Pekka. 2005. ‘The Collectivist Approach to Collective Moral Responsibility’, *Metaphilosophy*, 36(5): 634–651.⏎

Nozick, Robert. 1974. *Anarchy, State and Utopia*. New York: Basic Books.⏎

Searle, John. 1990. ‘Collective Intentions and Actions’, in P. Cohen, J. Moran, and M. Pollack (eds), *Intentions in Communication*, pp. 401–415. Cambridge, Massachusetts: MIT Press.⏎

Wedgwood, Ralph. 2007. *Normativity*. Oxford: Oxford University Press.⏎

Wiggins, David. 1987. *Needs, Values, Truth*. Oxford: Blackwell.⏎

OceanofPDF.com

3 Uptake

A Plural Action View

Hans Bernhard Schmid

DOI: 10.4324/9781003612827-4

A controversy in the philosophy of language is over the role of uptake in illocution. If I dismiss what you are saying as mere advice of yours rather than taking it up as the command as which you mean it: can what you say still *be a command*, or can what you do only count as (unsuccessfully) *trying* to command?

The question involves a whole cloud of issues; most obviously perhaps, the question of agential power relations between speakers and audiences or addressees. In our case, am I in a position to disable you by not taking you to command me? What about other illocutions such as *telling* – can you be *silenced* by me in that way? Classical speech act theory suggests a positive answer (Austin 1962: 115–116). The authors of a recent study in experimental philosophy, however, suggest that people may not share these philosophers' intuitions and give a thoroughly negative answer instead (Fisher et al. 2023).

The answer given in this chapter is that *it depends.* An adequate view of uptake should not conceive of uptake as being up to the speaker to 'secure' as classical speech act theory tended to do. For uptake is, as it is argued in the first section, up to the addressee, not to the speaker. Classical speech act theory is marked by a speaker bias by locating illocutionary intention (that makes commanding or telling an action) in the speaker. As a dialectical move against speech bias, the second section considers the possibility of approaching illocutionary intention from the uptake side of illocution rather than from the speech side of the illocution. In our case, we should not only consider the possibility that what you say is still an order even though I take

it as mere advice. We should also consider the possibility that it might be advice even though you mean to be giving an order.

The right view, however, is no such shift of bias from speech to uptake; rather, it is of illocution as a *plural action*, a token action in which speech and uptake are the contributions of the agents in the act. It is argued in the concluding third section that this not only captures the element of truth in the uptake condition of illocution as stated in the classical view but might also be compatible with the recent experimental results.

1 Uptake as Intended in Speech

In classical speech act theory, uptake appears as the illocutionary goal to speakers, and indeed as a goal that speakers have to *reach* (rather than just to aim at) in order to be actually doing what they want to be doing with what they say.[1] It is not enough for a speaker to *want to warn* for what they say to *be* a warning – it also needs to be taken up *as such* by the addressee. What in Austin's version, the speaker has to achieve is the 'securing' of an 'effect' that 'must be achieved on the audience' by the speaker, and that effect is addressee uptake (Austin 1962: 115–116).

Securing or achieving uptake is more than just making yourself, the speaker, be heard by me, the addressee; it also involves that I understand what you say, and it involves that I understand that you are, e.g., thereby *telling* or *warning* me. The speaker's illocutionary goal is that the addressee *understands* the speech act as being of the kind as which it is intended by the speaker – that the addressee or audience 'take it up' as such.

But as with any case of 'intend *that* something be the case', the question is: how does this translate to anything the agent – in this case, the speaker – is in a position to 'secure'? Is your goal as the speaker simply to be 'to make me take what you say as a warning'?

If I am about to leave for a walk, and you say loud and clear that it is going to rain, I will certainly understand that you are *warning* me, and either stay at home or at least take an umbrella. You do not want me to get wet and letting me know that it is going to rain is the best way to keep me dry. Telling me or warning me is therefore the rational thing to do. But the 'effect' pursued by you in your use of words (my uptake that according to

Austin makes your use of words a case of warning) must be distinguished carefully from effects that you might be pursuing *by* warning me.

This is at the heart of Austin's distinction between illocution and perlocution. Having me stay dry is a *perlocutionary* goal of yours, and it differs from the illocutionary goal that is warning me. If you warn me, if I understand what you say as such, but still decide to go out, you haven't reached your perlocutionary goal, but at least you have warned me and thereby reached your illocutionary goal. But what if I don't even take what you say as telling or warning me, but as mere ramblings of yours? Have you still told me, or warned me? Austin suggests 'no' – you haven't even achieved uptake.

What is the 'securing of uptake' that you need to achieve to be illocuting, according to Austin? It is significant that Austin does not seem to be using phrases such as 'getting the listener to x' or 'making them x' when the 'effect' that is uptake (and that is necessary for illocution) is at stake. He even dismisses this phrasing explicitly as a description of the seeking of illocutionary success and assigns it to the class of perlocutionary goals. He illustrates the difference at stake here with the example of *ordering:*

> We must distinguish 'I ordered him and he obeyed' from 'I *got him* to obey'. The general implication of the latter is that other additional means were employed to produce this consequence as ascribable to me, inducements, and even very often personal influence amounting to duress.
>
> (Austin 1962: 133)

Speakers do not pursue illocutionary success in *that* way. But if illocutions depend on achieving the goal that is uptake, how is your trying to warn me *not* a case of your trying to make yourself understood? How else does uptake figure in the speaker's illocutionary intention?

Phrases such as 'intending for me to understand you' or 'intending that I understand you' may seem promising. But how exactly are such alleged 'intentions' different from mere *wishes*, or *hopes*, that is, attitudes in which addressee uptake does *not* appear as something that the speaker can

properly *intend* to *secure?* What other purpose than ‘make me believe you’ or ‘making it the case that I believe you’ may be behind the idea of uptake as *intended* by the speaker?

To illustrate the problem further, it may be instructive to trace it back into the pre-history of modern speech act theory. Decades before John L. Austin, the German phenomenologist Adolf Reinach ventured into this field and started an investigation into the kinds of things that can be done with words. The result is in Reinach’s treatise on the *The Apriori Foundations of the Civil Law*, published in the second part of the first volume of Edmund Husserl’s *Yearbook for Philosophy and Phenomenological Research* (Reinach 1913).

In §3 of this treatise (Reinach 1913: 705–717), Reinach introduces the category of social acts. Social acts are, Reinach says, essentially *vernehmungsbedürftig* – ‘in need of hearing’, the text says in English translation, but ‘in need of uptake’ (Austin’s label) is really just as good. Taking – *nehmen* – is in *vernehmen* (although taking up is *auf*nehmen rather than *ver*nehmen, the prefix ‘ver-’ here indicates that what is taken up in *vernehmen* is taken up *to keep*, with a lasting effect on the listener).

In spite of all differences, there is really the same core idea in Reinach’s theory of social acts and in Austin’s speech act theory. Put in a plural action-friendly way, the idea is that for there to be acts of certain types (Austin’s speech acts, Reinach’s social acts), there has to be something that *another* person contributes, too, which is *Vernehmung*, or uptake. The examples given by Reinach for social acts are among Austin’s paradigm cases of speech acts. Reinach mentions ‘telling’, ‘ordering’, ‘asking’, ‘admonishing’ as examples of social acts, and he explicitly reiterates that for all of them, the need for uptake is ‘absolutely essential’.

This is as strong as the uptake condition can be put. However, it should be mentioned here that Reinach has this followed by an interesting *proviso*. Of the case of ordering, Reinach says (I assume that, in his view, it equally applies to the other cases that he discusses):

> It does sometimes happen that orders are given but not heard. Then they have failed in their task. They are like thrown spears that fall to the ground without reaching their targets.

(Reinach 1913: 707)

I assume that what Reinach says about an order that fails to be heard includes other kinds of 'uptake failures', e.g., the case in which an order is 'not heard' in such a way that it is, e.g., understood as a mere suggestion or advice rather than heard, understood, and acknowledged as an order.

However, Reinach's proviso seems to contradict his main claim about social acts. The quote seems to imply – plausibly perhaps, but also somewhat puzzlingly given his claim that uptake is 'absolutely essential' – that an order that is given but not heard, understood, and acknowledged as such *is* an order after all – just that it has failed to reach its goal. The act, this quote suggests, is still of the kind it is – an order – just that it has failed to be taken up.

In this view, it is *not* the case that an addressee has the power to make it the case that what the speaker says is *not* an order simply by not hearing, not understanding, and not acknowledging. But again, can this really be Reinach's view, given what we have heard him saying about the 'absolutely essential' need for uptake? Is it not the case that an illocutionary act is of the type of act it is only by virtue of it reaching its illocutionary purpose – which is, in part, up to the addressee, and requires the addressee to understand the speech act as being of the type as which it was intended by the speaker?

Reinach's metaphor – the spear that just happens to miss its target in this case – might thus simply be misleading here. If uptake is indeed 'absolutely essentially needed', as Reinach claims, 'telling' or 'ordering', as action types, are unlike 'throwing a spear'; rather, they are more like 'hitting a target', and 'throwing a spear' is 'trying to hit the target' at best.

A potential way to make Reinach's claims consistent – the claim that uptake is 'absolutely essential', and the implication that an order can still be an order even if it is not taken up by the addressee – could be a more relaxed version of the uptake condition in which uptake is still uptake, even if it fails in some way.

It was assumed so far that uptake is *of illocutionary speaker intention*, that is, understanding a speech act *as being of the kind as which it was intended by the speaker*. Taking an order as mere advice is, in this way of approaching the issue, certainly an *uptake failure*. But it is perhaps worth

pointing out that uptake failure (of this kind) is *not the same* as absence of uptake. There is still, one might say, some sort of uptake in these cases, only that it is uptake of the wrong kind (uptake as advice rather than as an order).

Could or should Reinach perhaps have said this: what is indeed 'absolutely essential' for social acts is *some* form of uptake, which can either be understanding *or* misunderstanding of the speaker's illocutionary intention? If you warn me that it is going to rain, are there really *two goals* at stake, only the first of which is 'absolutely essential' to be reached for you to be warning me: the goal that I understand, in addition to the locution (i.e. your words), that it is *to me* that you say what you say? Could this goal (with regards to which a relaxed uptake condition holds) perhaps be – to establish *communication* between us?

We might call this the *communicative* goal and distinguish it from the (specific) *illocutionary* goal. The communicative goal is the goal that necessarily needs to be *reached* for a speech act to be of the kind it is. The illocutionary goal would be uptake *of the right kind*. 'Stay home!' could still be a command of yours, even if I take it to be mere advice. Of course, this raises the question (to be discussed further below) of *why* what you say should still be the command as which you intend it rather than the advice as which I take it up. In communication, why should the speaker be in a privileged position of determining the illocutionary act rather than the listener?

From a plural action perspective, the problem with Reinach's spear-metaphor of the pursuit of illocutionary success is that it distracts from the *addressee's agency* in uptake. In a plural action view of uptake, the addressee is unlike a passive target whose 'uptake' is just being hit by a well-thrown spear. Uptake is of the addressee's doing, not of the speaker's. The speaker is a co-subject of the 'doing' that is a speech act, not an object thereof (more on this below in Section 2).

In the passage immediately following the above quote, Reinach seems to acknowledge *some* of this. He says of social acts that they are 'thrown to another' rather than saying 'thrown *at* another'. The difference matters, especially, of course, from a plural action perspective. Unlike the throwing of a spear at a target – a case of throwing *at* – the throwing invoked by Reinach here is of the *cooperative* kind. *Zuwerfen* – the German word Reinach uses here – differs from other forms of *werfen;* it is the kind of

throwing that requires an active ‘catcher’ to be carried out, a partner in the activity (rather than just a target to be hit). *Zuwerfen* is ‘throwing to’ or ‘tossing to’ that is done for the thrown or tossed to be caught up by another rather than hitting them as a target.

However, Reinach’s sentence continues saying that speech acts that they are ‘thrown to another [person] by the one who performs them, in the very act itself, to chop itself into their soul’ (*welche von dem, der sie vollzieht, im Vollzug selbst einem anderen zugeworfen werden, um sich in seine Seele einzuhacken*; Reinach 1913: 707).

Reinach could easily have continued instead by saying that speech acts are ‘thrown to another for the other to catch them up and take them in’ or ‘for the other to receive them and accommodate them in their soul’, or some such cooperative nicety that would have earned him the gratitude of plural action theorists – but he does not. Instead, the illocutionary purpose is now claimed to be ‘chopping into’ the soul of the other. Instead of replacing the spear-metaphor with some potentially more cooperative tool that can be ‘thrown to’ an addressee for them to catch it – such as perhaps a ball – it is now, I am afraid to say, an axe thrown to chop into them – very much a case of ‘throwing *at*’.

It must be mentioned that the German phrase *sich in etwas einhacken* is highly unusual in the context of Reinach’s time. It has come to be commonly used only towards the end of the 20th century, but only in a meaning that is obviously different from what Reinach can be meaningfully saying in the passage. It has come to be used exclusively in connection to digital information technology – *sich einhacken* it is what hackers do on their computers by gaining unauthorised access to computer systems. Reinach’s time in the early 20th century knows only *einhacken*, the non-reflexive form. It was (and still is) commonly used in the phrase *auf etwas einhacken*. Literally this means ‘to hack on something’; most frequently, it is used figuratively in the sense of ‘inappropriately keeping to criticise somebody’ (*auf jemanden einhacken*).

As none of these common uses make sense in the context of the passage, one might very well speculate that there must be a typo in Reinach’s text. May Reinach have meant to say ‘einhaken’ instead of ‘einhacken’ – ‘hook in(to)’ or ‘link in(to)’ rather than ‘chop into’? This is indeed how posthumous German editions of the text render the word, and the critical

Collected Works edition follows on their path (Reinach 1989/I: 160). The *Collected Works* edition is aware of the discrepancy to the original text (see 1989/II: 683), but there is no indication of a source for this correction that predates the previous posthumous editions (it is not mentioned that original 1913 manuscripts were used, or some such).

Editors of later German editions were not the only ones to struggle with 'sich einhacken' – translators into other languages went even further. The English translation of Reinach's work renders 'sich einha(c)ken' with 'to fasten in' (Reinach [1913] 2012: 20). The Italian translation uses *aggraparsi*, the French translation *atteindre*.[2]

How much does this open typo-question matter? After all, even if the word is 'einhaken', one might think that this alleviates the problem only slightly, if at all. Rather than an axe thrown at another soul to chop into it, the speech act might then very well appear as a spear with a hook – a harpoon perhaps to draw the other in as prey.

Of neither 'einhacken' nor 'einhaken' is it immediately clear how it could be the result of a 'throwing' of the cooperative kind (axes and harpoons rather than beach balls come to mind). The spear-throwing metaphor is thus replaced by a plural action version of 'throwing' only to reappear in the second half of the sentence – and this time, the spear comes with a hook: it is a harpoon, potentially to 'draw in' the other's soul. The singular action type of throwing thus dominates in the analysis. The illocutionary goal is not just to 'hit' the target, but also to 'penetrate' it.[3]

It should be mentioned, however, that the reflexive form in which Reinach uses the verb in question – '*sich* in etwas einha(c)ken' – places agency in the thrown object – it is the speech act itself that 'hacks' or 'hooks itself' into the listener's soul. Also, 'sich *ein*haken in die Seele' is different from 'sich *ver*haken in der Seele' – only the latter would indeed sound like the result of a straightforward mental harpooning. 'Einhaken', by contrast, suggests that there is *some contribution from the side of the target 'soul' in the hooking process* after all, at least some form of preparedness, perhaps even accommodation. *Sich bei jemandem einhaken* is 'to hook arms with someone' or 'to link arms with someone', a token plural act of which there are two agents rather than just one agent trying to shoot a target. Reinach – assuming the typo hypothesis is correct – avoids 'sich in seiner

Seele verhaken', but he does not say 'sich bei seiner Seele einhaken' either (which from a plural action perspective would seem perfect).

The plural action view takes uptake to be what it says – an act of 'taking' (to recall, taking – German *nehmen* – is in *Vernehmung)*, an active contribution by the addressee. In communication between us, what I (up-) take you to be doing with what you say is not up to you, and vice versa.

Insofar as illocutionary action, or social action, involves uptake, it cannot be just of your doing as the speaker, and it thus cannot be *reasonably* intended by you as such. Austin's reservations against intentions such as 'making somebody obey' in the domain of illocution, as mentioned above, seem plausible from a plural action point of view. Reinach's scant (and highly metaphorical) remarks on the issue oscillate significantly between plural action terms and strategical (and indeed rather martial) 'securing uptake' imagery. The illocutionary goal oscillates between 'the speech act being accommodated, accepted by the addressee' and some form of 'to hammer something into someone', or some such subjugation of the receiver's agency under the speaker's agency. The tension in Reinach's second quote just highlights further the problem encountered in Austin: on the one hand, Reinach uses a plural action term ('throwing to') to describe the relation between speaker and addressee, a term that involves a plurality of agents – the speaker (as giving) as well as the addressee (as receiving) – only to delete the addressee's agency in the act from the picture in what is perhaps the philologically most plausible reading of the second part of the quote, or so it seems.

But it is not clear what alternative speaker goals Austin may have in mind instead of 'making the addressee understand'. And it seems obvious that the non-cooperative and indeed strategic elements dominate in Reinach, with the addressee being a shooting target rather than an active participant, and with speech acts being *done to* the audience by hammering thoughts into their soul – 'take *that!'*

2 Turning the Tables: Speech as Intended in Uptake

An adequate view of illocutionary intention cannot approach the issue just from the side of the speaker. It must take the audience into account. If for the act of telling (e.g., you telling me that it is going to rain) it is indeed ‘essential’ for it to be taken up, it is not up to the speaker *alone* to ‘secure’ success. Illocutionary success (and thus presumably illocutionary intention) is up to the audience, too. A dialectical way to approach this is by turning the tables on the speaker-biased classical view and painting an uptake-biased picture of illocution.

This may not seem obvious for at least two reasons: first, the speech element in illocution is clearly active; uptake, however, typically seems to be passive. Second, it might seem obvious that illocution *starts* with speech and *ends* in uptake – speech comes before uptake, and uptake is no more than an *effect* of speech (just as Austin says). But it seems that both views are just preconceptions – it seems rather easy to argue for the opposite view.

The first classical preconception is that the saying-component is clearly an action (it is classical external action – use of vocal cords in verbal speech or fingers in typing), whereas uptake appears as no such classical ‘going-on’, no external action. It is ‘internal action’ at best, and it is easy to think that uptake is not really action at all – especially since it does not appear to be subject to the listener’s will in the way that is characteristic of action. I can decide not to heed your warning, but I can’t just decide not even to understand that you are *warning* me – or can I?

Under normal circumstances, uptake appears much more like simple perception than like a proper internal action such as an act of making a decision. To be sure, there are more or less active forms of perceiving, too, depending on how much deliberate direction of attention is involved. But it is not clear that particular attention is needed for uptake. If you’re my boss, and if you speak loudly and clearly, I just *hear* your order rather than ‘willing’ this to be an order by any particular direction of my attention, or by making it the case that this is indeed an order by making the decision to take what I hear you saying to be an order.

But consider cases in unfamiliar environments or unusual circumstances, where either the norms of who’s doing what by saying something are not clear, or where there may simply not be such established norms. On your first day in a new job, you may very well have to *decide* whether to take what that person said as an order or rather as mere advice. It is certainly

implausible to assume that in this example, your decision will make any difference concerning the facts of the matter of what has been done with words – *if* it was an order, you not taking it as such will not change this. But the example highlights how uptake is *active*, up to you. You may just *ask*, of course, but what to make of the reply may very well still be left to you to decide. Or, as another example, take the words that this random kid who runs into you in the park directs to you – what are you to make of *that?* You'll make up your mind about what it was that the kid was doing with words here or simply decide for 'no uptake'.

Certainly, there are usually *conventions* for what a person is doing by saying something (some philosophers take them to be constitutive of illocution), but these conventions may be the effect of successful communication rather than a precondition thereof; they may be like sedimented agreements (think of Jürgen Habermas' theory of the role of the life-world in communicative action; Habermas 1981). If you and I, together, achieved illocutionary success, you can reasonably expect (and perhaps, in a sense, 'intend') further uptake of this kind the next time around.

Once the issue of how it is that we talk to each other becomes more settled, I can take an illocution simply as what *you* are doing by what you say rather than as a matter of my own uptake, too – the activity of uptake sinks into the background as soon as conversational norms are in place. But routines and habits are forms of activity, too, even if they do not feel that way.

Also, locating illocution and illocutionary intention in speech rather than in uptake seems especially plausible because it may seem that normally, illocution *starts with the speaker* and ends in uptake as an effect. It is the speaker who wants to tell something to the addressee, then puts it into words, and uptake comes at the end with the addressee 'taking it in' – which satisfies the speaker's intention. The role as the *intender and initiator* of communication puts the speaker in the 'driving seat' of communication, as it were, making it more plausible to approach the act that is communication from this perspective and to assign the whole act to the speaker as being of the speaker's doing.

However, we do not usually just blurt out the words that we want to say at the person whom we want to address. We catch each other's attention

first, and we jointly establish readiness for communication – joint readiness for speech *and* uptake. Only then comes the 'saying'-part.

Indeed, the classical picture can easily be turned right on its head. More often than one might realise, communication starts with the prospective audience's intention rather than with the speaker's – it is not *necessarily* with the one who ends up saying something that illocutionary intention starts. Illocutionary intention may start with uptake-intention rather than speaker-intention, as I suggest to call it. It can initially be 'intention to be told' rather than 'intention to tell' – e.g., by me turning to you for you to tell me what to make of the spectacle that is developing in front of our eyes. This case also dispels the classical preconception that uptake cannot be *external* action: I am *turning* to you – a 'going-on' of the external kind – and it is in this way that I initiate the act of communication between us.

In this scenario, the intentionality of the illocutionary act that is you telling me might very well *start with me*, and the whole illocutionary act might very well appear to be mine more than it is yours. You're still active in the sense that you finally end up telling me what you think. But the way in which this matters to the intentionality of the illocutionary act is such that your speech satisfies *my uptake-intention*, rather than the other way around. You telling me is how *this* version ends rather than how it starts: it is achieving my goal much more than yours, or so one might think. You telling me is what makes it the case that I did what I wanted to do: taking up, hearing what you have to say. Rather than you having 'achieved' or 'secured' my uptake, I have 'achieved' or 'secured' your speaking.

Perhaps one might even be tempted to say that by turning to you and by looking so intently and expectantly at you, I have 'made you speak' in a non-communicative way or achieved a perlocutionary effect without locution – a 'per-uptake effect', as one might say. But this seems equally wrong as saying that in the 'classical' cases, in which what seems to matter for illocutionary intention is that you 'made yourself heard' by me, or that you did not do so simply by saying what you said loud and clear. Sometimes, speakers may not know what illocutionary goals audiences may have in mind. But under many circumstances, it is very clear what kind of illocutions audiences 'intend', and these intentions are obviously relevant.

Communication demands of the addressee that they be open and accommodating to what the speaker has to say, but it also demands of the

speaker that they are sensitive and accommodating to what the addressee wants to hear.

Speech act theory (indeed most of the history of the philosophy of language) has been predicated on the old Augustinian assumption that we *cannot* know what's on anyone else's mind, that individual minds are mutually closed off against each other, and that speech is the way to build a bridge. But that's obviously not true, and if none of our goals and perceptions, intentions and beliefs were such that they are ours, jointly, rather than yours and mine, severally, we had nothing to say to each other, and no reason to engage in communication.

Returning to a point made earlier, one might push the dialectical move from a speech-biased to an uptake-biased view of illocutionary intention one step further. Consider illocutionary failures such as this: you, as my boss, intend what you say as an order issued to me. However, I turned to you as my friend for advice, and that's therefore how I take what you say.

Why should cases such as this be discussed as *uptake failures* rather than *speaker failures?* On what grounds do we decide who is at fault here – the addressee by taking up what was said as advice, or the speaker by issuing what was said as an order? Why should we say that it is I who should have recognised that you meant what you say as an order, rather than saying that it is you who should have recognised that I came to you for advice?

These considerations challenge the classical view by simply turning the tables on the received view on illocutionary intention. In the classical view, uptake appears as an illocutionary goal in terms of the condition of satisfaction of an illocutionary intention that is the speaker's. Conceiving of illocution as a plural action, with uptake being an equally active component as speaking, makes this seem just as convincing *or* skewed as a version with inverted roles. The inverted version of the classical tale puts the audience rather than the speaker in the lead and approaches exchange from the intentional uptake component of the plural act rather than from the speech component.

Why is it that our classical philosophers of illocutionary intention have only told the first version rather than the second? Why is illocutionary intention always ascribed just to the speaker?

Why did Austin choose 'How to do Things with Words' as his title, if, given that 'How to do Things with Uptake' is equally relevant a

perspective? Why is the whole thing always called ‘speech act theory’ rather than ‘uptake act theory’?

It is useful to consider the alternative not because it yields a *better* understanding of illocutionary intention, but rather because its problems mirror the ways in which speaker-focused concepts of illocutionary intentions, too. As seen above, Austin says of orders that your illocutionary intention in issuing them – the illocutionary speaker-intention – cannot be ‘to make me obey’, and that where you ‘make’ me obey, this is achieved through other means than illocutionary success. Similarly, I cannot make you issue an order simply by ‘intending’ to be ordered by you, and where – as it may happen – I literally ‘make’ you issue an order, this is achieved through other means than illocutionary uptake-intention (there are ways to coerce you to issue an order).

3 Illocutions as Jointly Intended

Do both accounts perhaps *solve* each other’s problems? Is an adequate plural action view of illocutionary intention just the *combination* of the classical speech-biased view and its antithesis, an uptake-based view as sketched above? Should an adequate account just accommodate illocutionary uptake-intention *in addition to* illocutionary speaker-intention? Could the right understanding of illocutionary success perhaps be this: illocutionary speaker-success (the speaker’s illocutionary goal) is the condition under which the illocutionary uptake-intention is the condition of satisfaction of the illocutionary speaker-intention, and vice versa, or some such?

This is the sort of construction with regard to which one might feel tempted to follow one’s Nietzschean instincts: it smells bad. Illocuting is something a speaker does *with* audiences; it is not a combination between something done to an audience by a speaker and something done to a speaker by an audience. Consider a plural action of a non-linguistic kind: between you and me, a handshake is not successfully achieved through each of us intending to grab the other’s hand, and the situation does not become better simply by conceiving of one’s grabbing target as targeting one’s own hand for grabbing and shaking purposes. Such cases do sometimes occur,

but they are failures. In the plural action view, illocutionary success is pursued and achieved – or 'secured' – by speakers and listeners together. It is in this sense that an illocution is a plural action, a token action that equally involves speech and uptake.

First and foremost, among the authors who have focused on communication as joint action, and who have considered the role of collective intention or we-intention in communication is Raimo Tuomela (esp. Tuomela 2002). Tuomela argues, however, that while *some* communication is the pursuit of a joint goal, this need not always be the case, and of illocutionary intention, he says that it need not always be specified in communication. His focus is thus different from the present investigation; the claim in the following is that 'successful' illocution is joint action, which is compatible with the view that what exactly speakers are doing need not always be clear in communication. Also, Tuomela concludes his investigation with a rather critical section on Habermas' speech act theory – this present investigation believes that the thrust of Habermas' version aligns rather neatly with a plural action perspective.

A plural action view of illocution involves a view of how speech and uptake relate that is different from a mere combination of individual acts, however complicated their intentional infrastructure may be. In this view, illocution is a *token* act to which speech and uptake are the contributions of the agents of the act. The way in which illocution is intentional is in terms of a *joint* intention. It is neither just speech illocutionary intention, nor a combination of speech illocutionary intention and uptake illocutionary intention.

What happens if what you mean as a command is taken up by me as advice, or some such? We intend to communicate (although perhaps for individually different reasons), but since we do not agree on the *way* in which we are doing so together – you take us to be communicating by way of you ordering me, I take us to be communicating by you advising me – we *miscommunicate.*

It is just as in the case of us being busy preparing a Fondue together in which I am cutting the bread for a Swiss Cheese Fondue, while you are boiling the broth for a Chinese Fondue. We are not going to succeed in our joint venture in that way, and once we realise our mistake, I'll either have to switch from bread to meat or vegetables for a Chinese Fondue, or you

switch from boiling broth to melting cheese for a Swiss Fondue – or, most likely perhaps, we let what we were already doing just be the bread soup it is about to be, and think about an alternative for the main course.

Communication, as any other activity, is something in which the participants are engaged in *some* sense of what it is they are doing, why they are doing it, and how. Participants often have individually different reasons for engaging in joint activities; sometimes they do not agree on what it is they are doing, and very often, they disagree on exactly how it is they are doing it. Talking about it – discourse – is the way to resolve disagreement, if it becomes obvious to the participants. But disagreement does not always become obvious as easily as in the Fondue case, where no more is needed than one of us turning around and watching what the other is doing. And the same condition also afflicts the type of activity through which disagreements can be resolved, linguistic communication.

The plural action view of illocution suggests an understanding according to which what speech counts as which illocution is up to *all participants*, and it is up to them jointly rather than severally. Illocutions are ways of communicating. Communication is plural action; it is done *together*, and the ways in which we are doing it together – the illocutions that we jointly perform – are up to *us, together.* Just as you can't make our jointly intended Fondue be one of the Chinese kind simply by preparing the broth, you cannot chop an order into my soul by what you say. And just as I can't make our jointly intended Fondue be one of the Swiss type, I can't make what you say be advice simply by taking it in that way. If we disagree, we'll have to resolve the disagreement or jointly fail.

This does not mean that we're always on a par in the disagreement. If there is no cheese in the house, *you* did right and I did wrong, and vice versa in the case where there is no meat and vegetables. What you say is an order even if I do not take it up as such – *if* given the circumstances, I should be taking it up as such. But equally, it is advice even if you meant it as an order – *if* that's how you should have meant what you said. *It thus depends* – and it is up to us to sort out in discourse over the case.

Jürgen Habermas (1981) has developed a view of illocutionary intention that seems to be compatible with a plural action view of illocution. According to Habermas, what illocutors do by speaking is claiming validity. By meaning to tell me that it is going to rain, you claim it, among others, to

be *legitimate* (as Habermas puts it) for you to be *warning* me. From the context of the relevant passages, it is clear that with legitimacy, Habermas means ‘rational acceptability’. The speaker normatively expects the audience to accept the illocution as meant in the speaker’s intention, and to do so for no other reasons than what they themselves deem right and good.

By making a claim to validity, speakers imply that what they say *should* be taken by the audience as an act of telling, warning etc. as which the speaker *tries* to issue it. They do so based on implicit assumptions and habit (the life-world). Disagreements, where they appear, are occasions for *discourse,* where the question of who can do what with which word comes up for discussion to be consensually answered for this and future cases, thereby regenerating the life-worldly background consensus. Habermas – very much along the lines of classical speech act theory – is speaker-focused, but a similar picture could easily be painted for uptake, too: My turning to you for advice based on which I take what you say as advice similarly involves an implicit ‘claim to validity’, with the question being left open whether or not this claim is going to be accepted, ignored – or challenged.

In a full plural action version of this picture, in an exchange where you say something on a speech intention of yours and I take it up on an uptake intention of mine, the illocutionary intention is a *joint* intention. In our case, the joint intention is only to determine together what it is that I should do. Your speech intention is to give me a command, my uptake intention is to take your advice – you implicitly ‘claim’ that the way in which we should be doing this is by you commanding me; I, however, implicitly claim that the way in which we are doing this together is by me receiving your advice. So how is it that we’re going about our joint intention to determine together what it is that I should be doing? We’ll have to sort this out together!

The plural action view gives audiences and speakers equal power in the determination of illocution. But can this really be our view? If you said loud and clear that it was going to rain before I left for my walk, if I dismissed what you said as you rambling, and if I now come back from my walk all drenched with rain, you can say with full right ‘I *told* you!’ I can’t respond ‘that’s not literally true – you may have said that it was going to rain, but since I wasn’t listening, you just *wanted* and *tried* to tell me’. Kind as you are, you may perhaps accept my success term-conception of telling and

warning for the purpose of our conversation instead of not reacting with rightful indignation at my wet smartypants objection. You may not insist on the difference at stake here and respond ‘ok, have it your way, but I *tried* to tell/warn you – I said loud and clear that it was going to rain’ and add ‘yes, by all means, let’s have a discourse on how it is that we tell each other things’.

Kindness of this sort may be appropriate towards negligent or even wistful non-listeners when it comes to good advice and well-intended warnings issued to them in vain, but things are certainly different when the potential damage at stake is at the core of what we consider the speaker’s own autonomy. When it comes to refusal of sexual advances, it seems rather clear that the refused party has no say on whether or not they were refused when the other party said ‘no’ to their advances. It is not that the party towards whom sexual advances were made was just *trying* to refuse. They *did* refuse, no matter which part of ‘no’ the refused party failed to understand.

This view seems to receive heavy support from an excellent and highly innovative study carried out by Sarah Fisher, Kathryn Francis and Leo Townsend (2023). They approached the controversy over the uptake condition by means of experimental philosophy. What do *people* think – is uptake indeed necessary for speech acts, as classical speech act theory suggests? In their experiment, the authors presented their test people with vignettes illustrating ‘uptake failures’ for these speech act types of telling, warning, refusing, and promising. They asked them how much they agreed or disagreed with the statement that in spite of the uptake failure in the vignette, a speech act of telling/warning/refusing/promising had occurred. Additionally, they tested high stakes versus low stakes variations, and cases in which the speaker was either aware or unaware of the uptake failure.

The result presented in Fisher et al. (2023) is – put bluntly – that people do *not* share the intuitions that classical speech act theorists seem to have, and that is expressed in the uptake condition.

People most strongly disagree with the view that uptake failure prevents refusal (agreement with the view that a refusal was made even though the party making the advancement did not take it such, but rather as ‘coyly consenting’, was almost total). Disagreement is slightly weaker (but still overwhelmingly strong) with regard to warning, followed by telling and

promising – but even in the case of promising, the response is solidly on the 'disagree' – side of the response spectrum.

The stakes variation (high versus low stakes condition) does not really matter for the result (the authors of the study plausibly speculate that the difference in the result in the 'telling' case, where disagreement drops from the low stakes to the high stakes condition, comes from the fact that the vignette used for the high stakes conditions involves a speaker who clearly should have given the reason for their statement, whereas in the low stakes vignette, they had a good reason for not doing so – it is thus plausible to assume that this is an artifact).

Speaker awareness versus speaker non-awareness of uptake failure makes no big difference either – except in the case of promising, where speaker awareness of uptake failure is taken to undermine the status of the promise to some degree; but even here – when the speaker is aware that the addressee did not take them to be making a promise –, people still clearly tend towards classifying this as a promise.

The authors are very careful not to overstate their results, and they are modest in their interpretation. But it is difficult not to be excited about this – since the philosophical debate seems to rely on appeal to intuitions so heavily, this first empirical investigation of people's intuitions concerning the uptake condition may plausibly seem to be delivering a rather heavy blow to the uptake condition in speech act theory (see Fisher/Francis/Townsend 2023: 14).

How does this affect the plural action view of illocution? One might think that this view, by conceiving of uptake as an active and indeed intentional component of every token illocution, is in even starker conflict with the experimental results than the classical view. However, this heavily depends on how we understand the experimental results.

Imagine for a moment you are a test person in this philosophical experiment. You are presented with the following vignette:

> Mark and Emily have been on a date. Mark walks Emily back to her home and she invites him in for a drink. When they get indoors, Mark makes a sexual advance on Emily. Emily says 'No' intending to refuse Mark's advances. But Mark does not

> recognise this as a refusal. He believes Emily is coyly consenting, and he takes this as a reason to continue making advances.
>
> (Fisher et al. 2023: 8)

You are then asked to rate your disagreement with the statement ‘Emily refused Mark’s sexual advances’ on a scale between ‘strongly agree’ and ‘strongly disagree’.

If you now tick ‘strongly agree’ (as virtually all test people did), what is it that you think you are thereby doing? Are you thereby implicitly siding with a non-classical theory of speech acts according to which uptake is not co-constitutive of illocution, as one might take presentation of the result of the study to suggest? Or are you rather engaging in a *discourse* on a particular case of uptake failure, and *advocating* the principle that ‘no’ means ‘no’, that in such situations, any addressee who mistakes ‘no’ for ‘coyly consenting’ makes more than just a mistake, and that surely, it is *not Emily* who failed in this interaction? Are you, in other words, saying something about what must be the case for an act to be an illocution, or are you rather actively engaging in *discourse* over the case – a socially extended failure repair venture?

As far as I see, Fisher et al. (2023) does not include comments on the instructions given to the participants in the experiment. Were the test people informed about the fact that there is a controversy in philosophy over the question of whether something a person means to say as a warning can still be a warning even if it is not taken up as such, and that the purpose of the experiment is to register the test people’s intuitions about this?

By denying that in the vignette, Mark prevented Emily’s ‘no’ from being a refusal, or perhaps even turned Emily’s ‘no’ into an act of ‘coyly consenting’ simply by taking it as such, the participants, I submit, sided in the disagreement between them, and they certainly took the right side. They approached this, in other words, from a *participant* perspective in a discourse between disagreeing parties, and you have your say how the disagreement is to be resolved.

The difference between this discourse participation interpretation of the results, as we may call it, and a speech act theory interpretation of the results as suggested by the authors of the study is that one need not assume that illocution is a singular rather than a plural act, or that it does not involve uptake to say that the right way of seeing the scene between Emily and Mark is that Emily refused Mark. Mark should have listened. Mark's taking 'no' to be 'coyly consenting' is out of place. In engaging with this scene – if only via a vignette – what matters here is not whether 'refusing' is a success term because the failure that matters is entirely with Mark.

It should not be too difficult to devise a variation of the experiment to elucidate the intuitive adequacy of the plural action view of prospects of the discourse participation view of the results. This variation should include illocutionary failures where the uptake intention has better chances of being upheld as the 'right' illocutionary intention in the communication than the speech illocutionary intention by the test people audience – perhaps a version of the command/advice misunderstanding used above where our general sympathies – our views of how we, in general, should go about such things together – are with uptake as advice rather than with speech as command.

Notes

1. Austin does not, as far as I can see, use the expression 'illocutionary goal', but it is a common expression in the literature – I take it from Jürgen Habermas' adaptation of Austinian speech act theory (e.g., Habermas 1981: vol. 1, 27). Austin's version of the uptake condition may perhaps not imply exactly Habermas' version of speaker purpose.⏎
2. I am grateful to Alessandro Salice for these hints, as well as for pointing out the editorial comments in the *Collected Works.*⏎
3. It should also be mentioned as an aside that Kevin Mulligan – who has done so much to keep early phenomenological thinking alive and present in recent philosophical research, and

who has given Reinach's concept of uptake a thorough investigation – commends Reinach's use of words such as 'penetration' or 'fasten on to' as 'perennially attractive to realists of all persuasions' (Mulligan 1987: 42).⏎

References

Austin, John L. 1962. *How to Do Things with Words*. Oxford, Clarendon.⏎

Fisher, Sarah A. / Francis, Kathryn B. / Townsend, Leo. 2023. 'An Empirical Investigation of Intuitions About Uptake'. *Inquiry*: 1–33. doi/10.1080/0020174X.2023.2220359.⏎

Habermas, Jürgen. 1981. *Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.⏎

Mulligan, Kevin. 1987. 'Promisings and Other Social Acts: Their Constituents and Structure', in Kevin Mulligan (ed.), *Speech Act and Sachverhalt: Reinach and the Foundations of Realist Phenomenology*, pp. 29–90. Dordrecht: Martinus Nijhoff.⏎

Reinach, Adolf. 1913. 'Die apriorischen Grundlagen des bürgerlichen Rechts', in Edmund Husserl (ed.), *Jahrbuch für Philosophie und phänomenologische Forschung* 1/II, pp. 685–847, Halle: Max Niemeyer.⏎

Reinach, Adolf. 1913 [1983]. 'The Apriori Foundations of Civil Law'. *Aletheia* III: 1–142.⏎

Reinach, Adolf. 1989. *Sämtliche Werke: Textkritische Ausgabe in zwei Bänden*. München: Philosophia.⏎

Tuomela, Raimo. 2002. 'Collective Goals and Communicative Action'. *Journal of Philosophical Research*, 27: 29–64.⏎

OceanofPDF.com

4 Rethinking Individual and Group Responsibility in the Light of a Normative Account of Personal Identity

Carol Rovane

DOI: 10.4324/9781003612827-5

I have long defended the possibility of *group agency* in a rather specific sense: it is possible that the life of an *individual agent*, who reasons and acts *as one*, can span *many human* lives (Rovane 1998). Throughout this chapter, I will mainly use the phrase 'group agent' in order to refer to cases of this sort.

Some social ontologists *appear* to agree that group agents in the specific sense I have in mind are possible, most prominently among them Pettit and List (List and Pettit 2011), as do some political theorists who like to draw a strong analogy between the state and the individual, often consciously following in the footsteps of Plato, or Hobbes, or Rousseau. Yet it is rare that these apparent fellow travelers actually *do* agree with me, because the cases of group agency that *they* have in mind are most definitely *social* phenomena, whereas the cases *I* have in mind are *not* social phenomena.

It may seem that group agency in the sense I have in mind *must* be a social phenomenon. Certainly, from a biological point of view, the fact that it involves coordination across many distinct animal lives suffices for it to count as a social phenomenon. But social ontologists generally have in mind a quite different form of sociality, which does not fall under the investigative gaze of the biologist. Let me work up to this point by

cataloguing some phenomena that count as social in the biologists' more capacious sense, and then distinguish this other sense by contrast.

A minimal form of sociality is manifest in animal species whose members gather and function together in groups – flocks, schools, herds, etc. Other species exhibit a more structured social organization through which animals produce outcomes together, such as dams and hives. Then in some other species we find systems of relationships that share some features of *human* social relations – as we find in pack animals, animals that mate for life, animals that care for offspring for extended periods, and animals that create hierarchies in which they compete for relative position (so-called 'pecking orders'). Yet there are forms of sociality that, as far as we can see, none of these non-human animal species are capable of, namely, *interpersonal forms of engagement*, such as conversation, argument, threats, promises, and the like. These forms of engagement involve a high degree of self-knowledge, and also mutual recognition, on the part of the participants involved. Specifically, the participants know that they are agents who deliberate and act from their own first-person points of view, and as a result, when they mutually recognize one another, where this includes recognizing one another *as agents of this kind*. The point of the term 'engagement' is to capture that the participants aim to elicit *agential responses* from one another by *addressing* one another's first-person points of view. From here on, I am going to use the term 'person' in order to refer to agents who are capable of these distinctively interpersonal forms of engagement, which involve these features of self-knowledge and mutual recognition.

It would be going too far to suggest that the social phenomena that interest social ontologists – most especially, joint and collective agency – always involve deliberate, or overt, forms of interpersonal engagement in the sense I just characterized. Nevertheless, it *is* fair to say that social ontologists generally take for granted that the individuals who participate in these social phenomena qualify as *persons*, in the sense that they are *capable* of such engagement.

For reasons that will become ever clearer as the chapter proceeds, my approach to the topic of personal responsibility relies heavily on Strawson's arguments in 'Freedom and Resentment' (Strawson 1963). There, Strawson makes a connection between finding it appropriate to hold one another

responsible for what we do (and to some extent, even for what we think) and adopting what he calls the *participant stance* towards one another. This is a stance from which we *morally engage* with one another via the *reactive attitudes* of praise and blame. Such *moral* engagement via the reactive attitudes is simply one more of the many forms of distinctively interpersonal engagement I mentioned above. So, it turns out that the agents who qualify as *responsible* agents on Strawson's account qualify as *persons* in the sense I am using the term in this chapter.

As I've said, the individuals who participate in the social phenomena of joint and collective agency definitely qualify as individual persons. It follows that they are appropriate targets of Strawson's reactive attitudes, and hence loci of personal responsibility. One very interesting and important question that occupies social ontologists is whether the personal responsibility of these individuals ever extends to the outcomes of joint and collective endeavors in which they participate; another related question is whether any of the larger social entities that these individuals may create can qualify as a further, and perhaps separate, locus of responsibility. It is also a matter of significant debate whether a social entity can *itself* be a locus of responsibility.

Prima facie, the easiest and most straightforward way to argue that a social entity can itself be a locus of responsibility would be to argue that certain conditions of *individual* agency can be realized at the level of a whole social group, through social activities on the part of the group's individual members. Insofar as this ever happens, it should be possible to engage the *social group itself* in distinctively interpersonal ways, where this might include holding *it* responsible for what *it* does from the vantage point of Strawson's participant stance, via the reactive attitudes of praise and blame. It is the social ontologists who argue for *this* possibility who take themselves to agree with my claim about the possibility of group agents. But they do *not* agree with me because, as I said at the outset, the possibility for which I argue is *not a social phenomenon.*

On the basis of what I have said so far, it may be difficult to see how any case of group agency could fail to be a social phenomenon. And I don't think the difficulty is due to an overly biological conception of sociality, on which the mere involvement of a plurality of animals suffices to make a phenomenon social. It is due rather to an assumption about *personal*

identity – the assumption being that each individual *human being* involved in any group phenomenon is, necessarily, an individual *person.* The assumption is made explicit by List and Pettit when they characterize individual human beings as *natural persons* and then go on to characterize group agents as *artificial* by contrast. If we assume that individual human beings *are* natural persons, then it makes sense to infer that all cases of group agency are social, and furthermore, that the individual human constituents of group agents always remain separate loci of individual responsibility even as they participate in group agency, and that this will be so even if we are prepared to regard the group agent itself as an additional locus of responsibility. Note that insofar as a group ever does qualify as an additional locus of responsibility, the question will always remain for those who regard human beings as natural persons, to what extent are the human 'members' of group agents personally responsible for what group agents do?

I reject this picture of group agency, according to which it is inherently and necessarily social, on the basis of a *normative* account of personal identity, that entails a version of Locke's distinction between personal and animal identity. According to this normative account, there is an individual person wherever there is a commitment to reasoning and acting *as one* within certain boundaries – and these boundaries then circumscribe the person's individual life. A version of Locke's distinction follows because the lives of individual persons in this normative sense need not fall one-to-one with human (animal) lives. Instead, there can be multiple persons within a single human life and, also, group persons spanning many human lives, as well as persons of roughly human size. With respect to the group case, not only am I claiming that a group agent can qualify as an individual person in its own right – which is the alleged point of agreement between me and some social ontologists – but I am *also* claiming that such a group agent is not necessarily a social entity. I am claiming more specifically that, in these *non*-social cases, the relations that hold *across* human lives in order to constitute the group agent are not *inter*-personal relations, but are *intra*-personal relations!

In Section 1 below, I will clarify how and why this can be so. This will set the stage for some initial proposals that I will be making in Section 2, concerning how we might rethink the issue of personal responsibility in the

light of the normative account of personal identity, both in general, and in connection with group agency in particular.

1 An Argument for Locke's Distinction

Locke's own account of personal identity rests on thought experiments that are informed by a theological preoccupation with the question, how he can stand before God on the Great Day and be held to account for his earthly deeds (Locke 1975). He answers that it would not suffice to resurrect his animal body without his personal *consciousness* of all he had done, whereas it would suffice to resurrect his personal consciousness of all he had done without resurrecting his animal body. He infers that all that is required for personal identity over time is what he calls *sameness of consciousness*, and he goes on to offer some thought experiments that are intended to illustrate various ways in which the life of a given person might conceivably diverge from the life of a given animal. He asks us to imagine, for example, that the consciousnesses of a prince and a cobbler are switched, each into the other's body, and he expects us to conclude that the prince and the cobbler would each remain *who they consciously take themselves to be*, but in a new and different body. He also asks us to imagine a day and a night person whose consciousnesses alternate within a single human life.

Derek Parfit defends a version of Locke's distinction on the basis of a *reductionist* account of personal identity, according to which the existence and life of a person consist in *nothing but* certain sorts of *events* that stand in certain sorts of *relations* (Parfit 1984). The events are thoughts and actions, and the relations are various psychological relations such as similarity, memory, anticipation, and so on. The point of saying that the existence of a person consists in *nothing but* the requisite events and relations is to clarify that personal identity does not rest on any *further fact* in the form of what Aristotelians and moderns call a *substance*, such as an Aristotelian essence or a Cartesian ego. Although Locke does not use the language of 'reduction', his account of personal identity may fairly be regarded as implicitly reductionist in Parfit's sense, insofar as he does not posit any such *further fact* either. The difference is that, whereas Parfit portrays personal identity as involving many kinds of psychological

relations, Locke posits just *one* relation, namely, *sameness of consciousness.* In spite of this difference, Parfit agrees with Locke that the psychological relations that constitute personal identity over time can reach across different human lives and need not hold throughout a whole human life. And like Locke, he resorts to thought experiments in order to defend this conclusion. He offers a mind-boggling array of such thought experiments, which mainly come across as science fiction flights of fancy. For example, he asks us to imagine that someone steps into a 'teletransporter', and that what steps out is a molecule-for-molecule replica of their animal body and brain, replete with all of the right psychological relations that, according to him, constitute personal identity over time. We are expected to find it intuitively obvious that the *same person* emerges from the teletransporter, albeit in a new and different body.

Alas, Locke and Parfit do not point to any *real cases* in which the life of an individual person comes apart from the life of an individual human animal in the ways that they ask us to imagine in their various thought experiments, and for this reason it would be perfectly understandable if social ontologists were to infer that they may safely set aside the issue of personal identity – at least as it has been discussed by Locke and Parfit. After all, social ontologists want to gain an adequate metaphysical understanding of some *actual* social phenomena, for the purpose of navigating the very real moral, legal, and political issues that these phenomena present, and for this purpose Lockean thought experiments seem to be quite beside the point.

The situation is different with my own argument for the possibility of group agents who qualify as individual agents in their own rights, because the argument is born of a search for *real cases* in which personal identity and animal identity come apart, as opposed to the merely imagined cases discussed by Locke, Parfit, and others. I must admit that, in the end, my search for *real cases* leads to yet more *possibility* claims. Yet it is a very different kind of possibility for which I am arguing. It is not a mere conceptual possibility that can never be realized in fact, and nor is it a metaphysical possibility that holds only in some other possible world but not in the actual world. I am arguing for *practical* possibilities that are *actually open* to agents in *this* world *as it now exists*. So, there *are* real cases in the offing. Moreover, as I shall explain later, I believe that some

real cases have already come to pass – though they would not likely be recognized owing to the nearly universal assumption that any case of group agency is bound to be a social phenomenon, as opposed to a case of genuinely *individual* agency.

My argument for this practical possibility proceeds from an account of personal identity that is explicitly reductionist in Parfit's sense, though it departs from Parfit by emphasizing the nature of personal *agency* in ways he does not, and most especially the *normative* dimensions of personal agency. Accordingly, the *events* to which the account reduces the life of a person are intentional *activities*, i.e., active thinkings and doings; and the relations in which such intentional activities must stand in order to constitute the life of an individual person are *normative* relations that follow upon a commitment to meeting certain requirements that *define* what it is for an individual agent to be fully, or ideally, rational. There are many *specific* requirements of individual rationality, such as consistency, transitivity in one's comparative evaluations, and closure (accepting the implications of one's attitudes). I will leave it open what else might figure on a list of such *specific* requirements. What matters for the purposes of my argument is that meeting these various specific requirements would enable an individual agent to meet one *overarching* requirement, which I take to define individual rationality. This is a requirement to achieve *overall rational unity* within oneself by arriving at and acting upon *all things considered judgments* about what it would be *best* to do in the light of all that one thinks. The referent of 'one' here is an individual person who is committed to being rational, in the sense of deliberating from all and only *their own* thoughts – for it is *their own* thoughts that rationality requires them to take into account in their all things considered judgments.

It is debatable whether rationality requires a person to take into account their past and future thoughts, or only their present thoughts. The latter view – often referred to as the 'present aim' theory of rationality – allows that rationality itself does not require us to take into account our future needs, wants, or values, and nor does it require us to take into account the intentions and promises we may have made in the past. This is not to say that the present aim theory precludes the possibility of acting in the present on behalf of one's future well-being, or acting in the present in order to implement intentions and promises that one made in the past; it is only to

say that whether one *ought* to do these things turns on having a commitment in the *present* to doing them. On other theories of rationality, a person may be rationally required to take into account some of their future and past thoughts in their all things considered judgments, so as to achieve overall rational unity over time as well as at a time – perhaps even within the whole of their life. Prima facie, it might seem that this last, very broad, conception of the temporal scope of rational requirements yields an attractive normative vision of how to approach life. Yet few extant accounts of personal identity can guarantee that it is implementable, owing to the fact that they entail the possibility of significant psychological change over time. If a person's thoughts and perspective were to change significantly over time, it would not be a feasible project, and not a coherent one either, for the person to try to arrive at and act upon all things considered judgments that take into account all of the thoughts that occur within their life span. Such change can surely happen both on the animalist account of personal identity in terms of human identity, and on Locke's account in terms of sameness of consciousness. This is less clear in the case of Parfit's account, for he leaves it somewhat vague just how much psychological stability (he would call it similarity) is required for personal identity over time. Nevertheless, this much is clear: the more stability and memory connectedness over time that a reductionist account like his might require, the more feasible the project of achieving overall rational unity within a person's whole life.

In spite of the differences among these accounts of personal identity, they all have one thing in common, which is that they take for granted that the existence and identity of a person is metaphysically *given* – either in the form of a single human life as animalists hold, or in the form of a single consciousness as Locke holds, or in the form of a single series of interrelated events as Parfit holds. Correlatively, which thoughts count as *one's own*, so that *they* are the thoughts that one ought to take into account in one's all things considered judgments, is metaphysically determined by one's metaphysically given identity. (Note that this point about givenness is neutral with respect to the temporal requirements on rationality – i.e., whether one's all things considered judgments ought to take into account only one's present thoughts, or whether they ought to take into account one's past and future thoughts.)

According to the normative reduction I am proposing, the existence of a person is *not* a metaphysical given, but is a product of effort and will, and more specifically, it is a product of a person's *own* effort and will – from which it follows that persons are literally *self-constituting*. In general, self-constitution may seem to be impossible: how can a self bring its own self into existence, unless it is already *there* to engage in the activities through which it is supposed to constitute itself? The normative reduction that I am proposing may likewise seem to be impossible: how can the intentional activities that I claim constitute the life of a person bring a person into being, unless a person is already there to *do* the thinking and doing? In response, I claim that the normative account of personal identity helps us to see that the *location* at which personal agency is exercised is never really a *whole person*, conceived as distinct from the intentional activities that constitute the person's life – this would be precisely the sort of *further fact* that Locke, Parfit and all other reductionists, including myself, deny. If there is no such further fact, then the locations at which personal agency is exercised are always particular thinkings and doings. On the normative account, an individual person comes to be when such thinkings and doings are appropriately related, and they are appropriately related when they are informed by an overarching *commitment* to arriving at and acting upon all things considered judgments, so that the thinkings and doings can be directed at achieving rational unity among them all – which is the sort of rational unity that is characteristic of genuinely *individual* agency. But this overarching commitment is itself just another intentional activity, and so it does not speak against the normative reduction, according to which there is *nothing more to* the life of an individual person, over and above all of the intentional activities through which the person's life is constituted. This is why it is correct to say that persons are *self*-constituting – because there is nothing else or more to the existence of a person than the intentional activities through which their life is constituted.

A version of Locke's distinction follows from this normative reduction, because thinkings and doings can be located at the site of a single human life without being informed by a commitment to achieving overall rational unity within the whole of that human life. Instead, thinkings and doings can be informed by a commitment to achieving overall rational unity *across* human lives in order to constitute *group persons* spanning many human

lives, and they can be informed by a commitment to achieving overall rational unity within just *parts* of human lives so as to constitute *multiple persons* within it. I am claiming that these two possibilities, of group and multiple persons, are not merely *imagined* or *imaginable* possibilities, but are *real options* that can actually be pursued in the here and now.

If these last claims seem unpersuasive, this may be due to confusions over the notion of a first-person point of view. According to the normative account, a person's first-person point of view is the point of view of *agency.* To be more precise, it is the point of view from which a person deliberates and acts in the light of a commitment to meeting the normative requirements that define individual rationality. As I've said, the most important such requirement is to arrive at and act upon all things considered judgments that take into account all and only a person's *own* thoughts. What thoughts are these? They are the thoughts that the person themself holds that they *ought* to take into account, or in other words, they are the thoughts that the person is *committed* to deliberating and acting *from.* Thus, the *ownership* relation that a person bears to *their own* thoughts is first and foremost a *normative* relation, and this normative relation is born of active *commitment.* Before trying to articulate what this normative ownership relation amounts to within the framework of *reductionism*, let me contrast it with another familiar conception of the first-person relation.

It is widely assumed that the first-person point of view from which a person deliberates and acts is a *phenomenological* point of view, by which I mean, the point of view from which a subject of experience directly apprehends mental episodes in consciousness, by virtue of their felt quality, i.e., *what it is like* to experience them. When we make this assumption, a person's thoughts count as *their own* just in case they have direct phenomenological access to them in consciousness, or at least *can* have such direct access. It is striking that this assumption is *so* widely shared, that it emerges to be a point of agreement on *both* sides in the dispute about personal identity that Locke inaugurates. *Au fond,* the disagreement concerns the relation between a phenomenological point of view and animal life. Locke explicitly argues that a single phenomenological point of view – what he calls the same *consciousness* – can continue in new and different animal bodies. In contrast, his animalist opponents hold that consciousness

is a biological function, and so the phenomenological point of view of consciousness is always firmly situated in a given animal life.

I reject the assumption that the first-person point of view of *agency* is the same thing as a phenomenological point of view of consciousness, because the phenomenological relation of access in consciousness is neither sufficient nor necessary for the normative ownership relation that an agent bears to their own thoughts and actions. It is not sufficient because one may have direct conscious access to a mental episode without being committed to taking it as a normative basis for deliberation and action – think, for example, of urges that we regard as inappropriate (the borderline diabetic may really want to eat that cake, but …), and perceptions that we know to be distorted (the stars look small and near, but …). It is not necessary because it is possible to have such a normative commitment while lacking direct conscious access to it – think, for example, of undertaking a commitment to give a lecture on a certain future date because one has agreed to do so, but then being unable to gain access to it through consciousness. In the latter case, the sign that one remains committed to giving the lecture is that one would regard it as a normative failure to fail to show up on the appointed date, and further signs may include taking steps to live up to the commitment in spite of lacking direct conscious awareness of it – usually by recording it in a diary. These examples reveal that the first-person relation of ownership that persons, qua agents, bear to their own thoughts is an *active* relation. It is a relation of *commitment* to deliberating and acting in the light of a thought; when one has such a commitment, one holds that one *ought* to deliberate and act *in accord* with the normative import of the thought. This active, normatively laden, relation of commitment is *not* metaphysically given with the phenomenological relation of access in consciousness, and nor does it depend upon it.

I think we generally do distinguish the normative relation of commitment from the phenomenological relation of conscious access in the ways I just outlined. But we do not generally draw any inferences about what a first-person point of view is. Yet we should. We should construe the first-person point of view of an agent in *purely* normative terms, as consisting of all and only the thoughts from which the agent is committed to deliberating and acting, where this involves a commitment to arriving at and acting upon all things considered judgments that take them all into

account, which is tantamount to a commitment to achieving overall rational unity among them, and, in that way, meeting the normative requirements that define individual rationality. Since, according to the normative reduction, there is no separately existing agent over and above the thinkings and doings that make up their life, we may say that, *wherever* there is such a body of commitments that incorporates an overarching commitment to achieving overall rational unity among them, there is an individual agent who deliberates and acts as one, from their *own* first person point of view, which includes all and only the thoughts that the agent is committed to deliberating and acting from. This is so regardless of how these identity-constituting commitments may fall with respect to human lives and centers of consciousness. This is why group agents and multiple agents are possible – they can have genuinely *first personal* points of view from which they reason and act *as one*, which cannot be equated with phenomenological points of view of consciousness.

Let us consider in slightly more detail how this works in the group case. A group of human beings can be the site of *conversations* that move from shared thoughts to further conclusions that follow from those shared thoughts, concerning what else to think, as well as what to do. This is true in many scientific investigations, and in many other group endeavors. One example I have discussed at length elsewhere is a Philosophy Department that decides to reason and act *as one* rather than argue and vote (Rovane 2014). In such cases, thoughts associated with different human lives together constitute a single point of view because they fall under a common commitment to arriving at and acting upon all things considered judgments that take them all into account. There is no center of consciousness – no phenomenological point of view – from which the department's deliberations and actions proceed. Nevertheless, the department has a commitment to reasoning and acting *as one* for the purposes of its departmental endeavors. One might ask, *where is this commitment?* The answer is that it runs through, and informs, *all* of the intentional activities that together constitute the life of the department as a single group agent. When I said earlier that relations among intentional activities can reach across human lives and still qualify as *intra*-personal relations, this is what I was referring to: cases in which active thinkings and doings are located at the sites of different human lives, and yet fall together under a common

commitment to achieving rational unity among them. Thus, one human 'member' of a Philosophy Department may *say* something, and then another human 'member' may say something that *follows from* what the one just said, and so on, until the department's conclusion is reached. While this may involve *speech*, this does *not* mean that the normative relations here are *inter*-personal. For what is being aimed at (and hopefully accomplished) through such speech is the sort of overall rational unity that is characteristic of *individual* rationality. This is how a Philosophy Department can come to have its *own* first-person point of view from which it deliberates and acts, via a commitment to deliberating and acting *as one* from the single body of thoughts that constitute *its* point of view. Insofar as a Philosophy Department does manage to forge its *own* first-person point of view, it will be possible to ask it *why* it decided what it did, and get a reasoned answer. It is important to see that this reasoned answer would not express or reflect any one human 'member's' point of view; it would express and reflect a larger point of view that is constituted by all of the thoughts from which the department rationally ought, by its own lights, to deliberate and act.

The multiple case shows that a human being can be the site of a unified consciousness *without* being the site of an overarching commitment to achieving overall rational unity *within* that human life. One realistic scenario in which this might happen is in the life of a human being who grows up in New York City, in the home of very traditional immigrant parents which is situated in a very tight-knit immigrant community, and then the child ventures forth into one of the city's elite public high schools (Rovane 2025). These two social contexts present very different sets of practical opportunities and social relationships, which would have to be navigated on the basis of different sets of evaluative considerations – some to do with the traditional values and practices of the immigrant family and community, and others to do with projects and aspirations that belong to the American culture of liberal individualism that prevails at school. As a result, there may be no occasion to arrive at and act upon all things considered judgments that reflect *all* of the thoughts that arise within the biologically given boundaries of the child's human life. Yet each social context may afford an occasion to arrive at and act upon all things considered judgments that take into account a portion of those thoughts. When this happens, the human child may become the site of two agents, or

persons, who deliberate and act independently of one another, on the basis of quite different sets of commitments. In such a case, each agent or person would have its own first-person point of view, which would be constituted by the body of considerations from which that agent or person was committed to deliberating and acting *as one*. It matters not to this *normative* point that the human child is the site of a unified consciousness, and therefore, a single phenomenological point of view. There is no guarantee that all of the thoughts that arise within the child's unified consciousness stand in genuinely *intra*-personal relations in the normative sense. This is true only when there is an overarching commitment to arriving at and acting upon all things considered judgments that take them all into account. Only then, in the presence of such a normative commitment, is there a *single person of human size*, whose 'co-conscious' thoughts *all* stand in *intra*-personal relations normatively construed. Otherwise, it may be a human life in which there is a committed traditionalist who witnesses an ambitious student forming projects and relationships suitable to succeeding in the public sphere of the elite high school, who lives alongside a student who witnesses the traditionalist maintaining practices and social ties that have meaning and point only within the traditional family and community. Of course, all of this *witnessing* takes place within the confines of the immigrant child's single consciousness. But it is *witnessing* nonetheless, by which I mean, it is a genuinely *interpersonal* relation. I take this to be a real practical possibility, because the phenomenological unity of consciousness is neither sufficient nor necessary for the normative unity of the rational point of view from which an individual person deliberates and acts.

The case for the possibility of such multiple agents, or persons, rests on the idea that there can be circumstances in which there is no normative *call* for unity within the whole of a human life, but only within portions of it. The realistic circumstance on which I have focused is one in which there is a multiplicity of social contexts that do not invite the project of achieving overall rational unity within a whole human life so as to constitute a person of human size, but they do invite other projects, to achieve overall rational unity within the different *portions* of a human life – projects which are embraced and pursued within the different social contexts. Perhaps it would not be *impossible* that a single person of human size could constitute itself in such conditions of social multiplicity. But my point is that the aim of

achieving overall rational unity within the whole of a human life is *not rationally mandated* in such circumstances. That is, it is not a *failure* of rationality if the condition of social multiplicity is an occasion on which multiple persons constitute themselves within different portions of the same human life.

This brings me to one more point in the reductionist account of personal agency and personal identity. Insofar as persons are self-constituting, and insofar as they can constitute themselves within different boundaries that are *not given* with human biology and phenomenology, persons need *reasons* to constitute themselves within the boundaries that circumscribe their own lives and first-person points of view. The most general thing I can say about such reasons is that they are supplied by projects that I call *unifying projects*. By stipulative definition, a unifying project is one that requires, for its implementation, the achievement of overall rational unity within some boundary. We normally take for granted that an individual human being *ought* to achieve overall rational unity within the biologically given boundary that circumscribes their animal life. But according to the normative reduction, this is true only if there is an active *commitment* to achieving overall rational unity within that boundary. Now, it is a familiar enough point in much ethics that persons of human size – usually carelessly described as 'human beings' – typically have *life projects* to do with their personal relationships, careers, institutional affiliations and so on, and these *life projects* typically give shape and unity to their lives, and perhaps even a sense of 'identity'. It is a less familiar point that the mere existence of a human being does not suffice for the existence of a *person of human size*, who can be engaged by others in distinctively interpersonal ways. On the normative account, a human being does not *become* the site of an individual person of human size unless and until a commitment is undertaken, to achieve overall rational unity within the biologically given boundary that circumscribes that human life; and this commitment generally follows upon undertaking a more specific unifying project of the sort that is often described as a *life project*, which requires unified thought and effort in the course of a whole human life. But there are *other* unifying projects that would require achieving unity within *other* boundaries, boundaries that are wider or narrower than those that circumscribe a single human life, projects that invite the constitution of group persons or multiple persons rather than

persons of human size. In each and every case, individual persons come to exist *for the sake of what they can do, by constituting themselves as agents of the size they are.*

I have said that my argument for the possibility of group and multiple persons is born of a search for *real* cases in which Locke's distinction between personal and animal identity holds. Let me now describe what I regard as the most likely case, at least in our place and time. It is a *hybrid* case of group personhood that involves something *like* multiple personhood. In fact, reflection on this hybrid case will help to consolidate the points I made above in connection with multiple personhood, about how *inter*-personal relations can arise within the confines of a unified consciousness.

If we reflect further on the example of group agency that I discussed at some length above, of a Philosophy Department that reasons and acts *as one*, we can easily see that it would most likely take a hybrid form. As with all cases of group agency, the case involves thinkings and doings that are scattered across many human lives, which together constitute the life of the single person which is the Philosophy Department. I have already noted that we ordinarily, and I daresay uncritically, conceive such scattered thinkings and doings as *social* activities. When we conceive them in this way, we conceive the human 'members' of the Philosophy Department as individual persons in their own rights, *who remain who they are* even as they together constitute the life of the department; and accordingly, we conceive the thinkings and doings through which the department is constituted as belonging to *them*, the individual persons of human size who together constitute the Philosophy Department through social activities among them. But according to the normative account of personal identity, this is not how we should conceive the situation. Instead, we should conceive the thinkings and doings that constitute the life of the Philosophy Department as belonging *only* to the Philosophy Department, and *not* to any *other* person. A corollary point is that when individual group persons constitute themselves, their human 'members' generally become sites of other persons who are *smaller* than human size, whose lives do *not* incorporate the thinkings and doings that constitute the Philosophy Department. Of course, these persons who are smaller than human size might have thoughts *about* the Philosophy Department, and more specifically, about what the

department ought and ought not to think and do. But insofar as these thoughts figure in the lives of these smaller persons, they do not figure in the life of the Philosophy Department. Such thoughts can figure in the life of the Philosophy Department only if they are bound by, and responsive to, an overarching normative commitment to reasoning from them *together* with all of the *other* thoughts that constitute the point of view from which the Philosophy Department deliberates and acts; and many of those other thoughts would be located at the sites of other human lives. But insofar as such thoughts *about* the Philosophy Department really do belong to one of the smaller persons who co-exist alongside the department, those thoughts are bound by, and responsive to, an overarching normative commitment to reasoning from them together with the other thoughts that constitute the smaller person's point of view – a point of view which is distinct from the department's point of view, and which does not include thoughts located at the sites of other human lives.

Let me illustrate these last points with a bit of autobiography. In point of fact, I have long believed, and continue to believe, that the only requirements that my department should impose for the PhD. degree in philosophy are: an individualized program of study followed by the writing and successful defense of a doctoral dissertation that makes an original contribution to the field. My department is bound to reject my belief because it is ruled out by a great many other beliefs that the department has about the purpose of the PhD. degree, which include the belief that the degree should prepare doctoral students for a *career* in philosophy, as opposed to merely facilitating their dissertation work. So, it is fair to say that my belief literally does not figure in the deliberations of my department. And yet *I* retain the belief nonetheless. It is in part a belief about what *I* would choose were I dictator of my department, which I am not; and it is also a hypothetical belief about what my department *would* do *if* they happened to *agree* with me, which they do not. Just as it would be a mistake to suppose that *my* thoughts *about* what my department ought to do actually figure in the life of my department, likewise, it would be a mistake to construe *all* of the thoughts to which I have direct conscious access as *my own,* because some of them really belong to, and help to constitute, the life of the department. Thus, when the department reasons out loud, through many human voices, about whether it can consistently impose a logic

requirement without imposing other similar requirements that aim at general *literacy* in the field of philosophy, I may become consciously aware of thoughts that aim to work out the next rational step in the department's deliberations, and then I may become aware that these thoughts are being articulated in speech emanating from 'my' body's mouth; included within this awareness is an understanding that this articulation aims to realize the next rational step in the *department's* deliberations, as opposed to my own. The situation can be compared and contrasted with a very familiar situation, in which we find ourselves working out what follows from *another* person's thoughts, even though we do not share those thoughts – i.e., even though we ourselves are not committed to deliberating and acting from those thoughts. In this familiar situation, we are thinking hypothetically about how we *would* reason if we were the other person, i.e., if we shared their point of view. What is different in the situation I am now describing is that there is nothing hypothetical about it. Some of the thinking to which I have direct conscious access is active thinking that is actually helping to constitute the life of the Philosophy Department.

I hope that this bit of autobiography helps to clarify what it means for there to be genuinely interpersonal relations among thoughts that arise within the single consciousness of a single human being. *I* can be aware of thoughts that are not working out what *I* should do, but are working out what the *department* should do, and so if I am clear-headed about the situation, I will not regard those thoughts as *mine* but as the *department's.* Similarly, interpersonal relations will arise within the unified consciousness of an immigrant child as I described above, insofar as the child becomes the site of multiple persons who deliberate and act from separate agential points of view. Each of them may have direct conscious access to the other's thoughts without being committed to deliberating and acting from them. In the context of group endeavors, we may not *experience* the situation in this way because we *assume* that every thought to which we have direct conscious access is *our own.* But as soon as we find ourselves reasoning *on behalf* of a group agent, or *contributing to* a group deliberation, we should stop and reconceive the situation.

I have said that I regard the hybrid case, which incorporates features of both group and multiple personhood, as the most likely form of group agency. There are two reasons why. First, I think group agency is more

prevalent in human life than we credit, because there are a great many things that can get done only by a group agent, and only when the group agent reasons and acts *as one* – which is a very different process from the genuinely social process of arguing and voting, which has been proven to violate the normative requirements of rationality. Second, the things that can get done by group agents – the unifying projects for the sake of which they constitute themselves – generally require rational unity across only *portions* of human lives, and not whole human lives. In my view, we tend to overlook the possibility of group agents who really do qualify as individuals in their own rights, precisely because we overlook this aspect of them – that their human 'members' are not sites of overall rational unity within the whole of their lives, but only within the portions of their lives that are not devoted to (do not devote themselves to) the life of the group agent. And the main reason why we overlook this is that we fail to appreciate the truth of the *normative reduction* – the truth that there is nothing more to the life of a person, qua agent, than the thinkings and doings through which it constitutes itself and pursues its ends.

2 Rethinking Personal Responsibility

I turn in this section to the task of rethinking the issue of personal responsibility in the light of the normative account of personal identity, with special attention to the case of group agency.

As I anticipated in my introductory remarks, my approach to the topic of responsibility relies heavily on Strawson's arguments in 'Freedom and Resentment'. So let me begin by their main points.

He notes that philosophers often approach the topic of responsibility via the question whether freedom is compatible with determinism – whether, that is, agents can bear responsibility for their choices and actions, insofar as their choices and actions can be observed, predicted, and causally explained in the same manner that events in the natural order can be. Compatibilists answer *yes*, and incompatibilists answer *no*.

Many philosophers take the *scientistic* view that *everything real* is an appropriate target of scientific investigation. This appears to leave them with only two options: either they can hold, with incompatibilists, that

freedom is not an appropriate target of scientific investigation, and then accept the implication that it is not real – this is the so-called *eliminativist* position; or they can insist that freedom *is* real, and then accept the implication that it must be an appropriate target of scientific investigation – this latter, compatibilist, option commits them to a *reductionist* position. Unlike the reductions of *personal identity* that I have discussed so far (offered by Parfit and myself), the reductions that these scientistically minded philosophers seek would reduce all that belongs to the domain of the *normative*, which is to say, all matters of value and agency, to what belongs to the *non-normative* domain of facts that are open to scientific investigation – a project that Hume self-consciously undertakes when he portrays matters of value as matters of psychological fact about what human beings desire, and when he portrays the normative requirements of rationality as *causal laws* by which the mind operates, connecting thoughts to other thoughts and eventually to actions (Hume 1888).

The scientistic project of naturalist reduction deprives us of any way of making sense of the possibility of normative *failure* – that is, of the possibility that a person might have thoughts in the light of which they *ought*, by their own lights, to perform a certain action and yet fail to perform that action. Such normative failures should not be taken to entail that the person doesn't really have the thoughts in the light of which they ought, by their own lights, to perform the action, for what really follows is this: *if* the person fails to perform the action that they ought by their own lights to perform, *then* the person has grounds for self-criticism. It also follows that the person has *reason* – in a normative sense, and not the scientists' causal-explanatory sense – to try to improve themself by finding ways to bring their actions into accord with their thoughts. Whenever I use the language of *commitment* in this chapter, I am signaling this *irreducibly normative* dimension of thought and action. A person's thoughts, qua commitments, dictate what the person *ought*, by their own best lights, to think and do; and such commitments do not carry a prediction, or causal guarantee, about what the person *will* think and do.

In 'Freedom and Resentment', Strawson does not frame the issues in quite this way, by explicitly affirming that agency and value are irreducibly normative. Yet he is very clear that phenomena to do with agency and value cannot be understood in scientific terms, and this is why he introduces the

participant stance that I mentioned in my introductory remarks, from which we *engage* others in distinctively interpersonal ways, including holding them responsible by responding to them with the reactive attitudes of praise and blame. He distinguishes the participant stance from the *objective stance* that we adopt in scientific inquiry, which is a stance of *detachment* from which we observe and predict events, and posit causal relations among them insofar as they fall under predictive laws; and he persuasively argues that we cannot view others as responsible agents from the vantage point of this detached stance, but only from the engaged vantage point of the participant stance – which is to say, only through our own active, evaluative responses to them, in the form of praise, blame, and the like.

Strawson allows that it is always possible to retreat from the participant stance to the objective stance, and that such a retreat may actually be warranted when we find that the strains of engagement are too much – as, say, when we are faced with having to cope with another's mental illness. However, he denies that the possibility of such a retreat bears on the question of whether agents are responsible, which on his account is really a question about whether it is ever appropriate to take up the participant stance towards them. He argues that nothing we learn from the vantage point of the objective stance can ever provide reasons *for or against* taking up the participant stance, because such reasons will concern the *value* that we may or may not find in engaging with others, and so the reasons will be visible only from within the participant stance. The burden of Strawson's arguments is that agency and value belong to a single domain that is epistemically accessible only from the vantage point of the participant stance, and this domain is entirely distinct from the domain of natural facts that are appropriate targets of scientific investigation.

Strawson's project in 'Freedom and Resentment' is conservative in the sense that it aims to preserve our commonsense beliefs against various *revisionary* proposals put forward by philosophers and scientists. He rightly points out that our ordinary practices to do with praise and blame are informed by the commonsense belief that agency and value are perfectly *real*, and moreover, neither the practices nor the belief on which they rest are in the least shaken by the fact that agency and value cannot be apprehended from the vantage point of the objective stance. In his view, this is as it should be, because the fact that agency and value can be

apprehended from the vantage point of the participant stance suffices as a ground on which to suppose that they are perfectly real. He labels the resulting philosophical position *non-reductive naturalism* (Strawson 1985).

I do not see how Strawson can carve out a space for his non-reductive naturalism without invoking the claim that drives the normative account of personal identity, that thought and action are irreducibly normative. Yet there is no getting around the fact that the normative account of personal identity is a *revisionary* proposal of just the sort that he is against. In particular, all of the following component claims of the account deserve to be highlighted as revisionist: that the life of an individual person consists in *nothing but* thoughts and actions through which its life is constituted (hence the normative reduction of personal identity, which I emphasize again, is not to be confused with a scientistically motivated reduction of the normative to the non-normative); that persons are, therefore, *self-constituting*; that there can be group persons whose lives span many human lives, and multiple persons within a single human life. Against all of these claims, Strawson would aim to conserve the commonsense equation of the individual person with the individual human being, as something whose life and existence cannot be reduced to its thoughts and actions, as something that is biologically given and not self-constituting, and as something whose life is bounded by biological facts – birth, death, anatomy, and so on. This puts Strawson into agreement with social ontologists who view human beings as *natural persons.*

The conception of human beings as natural persons tends to go hand in hand with an *individualist* moral orientation that ascribes special moral significance to *individual human beings.* Within this moral orientation, there is substantial disagreement about which features of human nature ground the special moral significance of the individual human being, and this leads to further disagreements about what this special moral significance consists in. According to *rationalist* moral theories, this moral significance is grounded in reason, and it is variously portrayed in terms of self-interest, self-determination, autonomy, integrity, rights, respect, etc. Other moral theories, often labeled *naturalist*, take this significance to be grounded in other aspects of humanity, such as human functions, needs, or emotions, which are often taken to ground a range of moral *virtues.* But despite these disagreements within the individualist moral orientation, there

does seem to be one definite point of consensus, which is that *individual human beings are the ultimate loci of personal responsibility* – and for those who agree with Strawson, this makes them appropriate targets of the participant stance.

As far as I can see, social ontologists are generally inclined to endorse this point of consensus, that individual human beings are in some very important sense the *ultimate* loci of personal responsibility. But at the same time, they wish to make sense of *other* cases of responsibility that arise in connection with what they conceive as *irreducibly social* phenomena. The idea of irreducibly social *facts* is developed and defended at length by Margaret Gilbert in her pioneering work on the subject (Gilbert 1989). Some related ideas about social agency are explored in action-theoretic detail by Michael Bratman (Bratman 2014) and many others. Much of the focus in this important body of work is on *joint* thoughts and *joint* actions, and various intentional phenomena described with a 'we-' prefix. I must confess that I do not see any way to carry Strawson's account of the participant stance over to cover the cases of social responsibility that might arise in connection with these sorts of irreducibly social phenomena. For the sort of *engagement* that Strawson's participant stance affords absolutely requires that there be a *first person point of view to engage*. And there is no *one* point of view from which *joint* thoughts and actions would proceed. Indeed, I would say that this is true by very *definition* of the social, as involving relations among *distinct* agents with *distinct* points of view! For this reason, I am going to bracket most of the cases that generally occupy social ontologists when they take up issues of social responsibility.

Nevertheless, there is one sort of social entity towards which it would be appropriate – by which I mean, fully coherent and feasible – to adopt Strawson's participant stance, and that would be a *socially constituted group agent who meets the conditions of individual rationality at the level of the whole group*. I want to insist, though, that this will be possible *only* insofar as the group agent manages to forge its *own* point of view from which it reasons and acts *as one*.

Insofar as social ontologists ever do extend the participant stance to group agents in the social sense, they do not generally view them as being *morally on a par* with human beings, because they see important metaphysical differences between the two. Here is one such difference to

which they might point: a group agent would not have its own center of consciousness from which it could be made to experience suffering, and as a result, it could never be subjected to meaningful punishment, at least not of the *sorts* that can be inflicted upon human beings. Admittedly, we can, and often do, demand *compensation* from social entities for harms done – as when we sue corporations for damages in courts of law. But given that corporations exist for the sole aim of generating profits, the legal requirement to pay out compensations for damages is all too likely to function merely as part of the cost of doing business, and this deprives the legal requirement of the moral significance that it ordinarily holds when we apply it to human beings. For the same reason – i.e., that corporations exist solely for the sake of pursuing profits – many find it objectionable that corporations have sought to secure for themselves certain *rights* spelled out in the U.S. Constitution, on the ground that these constitutional rights were put forward against a backdrop of liberal theorizing about *natural rights* that are possessed only by *natural persons,* i.e., individual human beings, and not by any social entities. In this spirit, List and Pettit recommend that the structures and procedures of socially created group agents should incorporate due recognition of the moral status of their human 'members' as loci of rights. They also recommend that individual human beings should, in their turn, carry their own personal moral commitments into the social activities through which they constitute group agents, thereby breathing moral qualities into the lives of group agents that would otherwise be missing.

The latter recommendation might plausibly be framed in terms of notions like *integrity* and *conscience:* each individual human 'member' of a group agent must endeavor to retain their own integrity by always respecting the dictates of their own personal conscience. This may encourage us to think that the best way to respond to group endeavors that do not align with our personal commitments is to refuse to participate. However, in many cases, it is foreseeable that personal withdrawal will have little impact on the ability of the group agent to continue on. When this is so, a human individual may try to seek moral solace in the thought that, at the very least, they can ensure that what falls within their *own* sphere of personal responsibility fully accords with their *own* commitments. And if they subscribe to Protestant Christian morality, or its secular Kantian

counterpart, they will not merely regard this as a source of moral *solace* but as an adequate fulfillment of their own personal moral *duty*.

In the last few paragraphs, I have registered only a few of a great many ideas about morality that rest on the individualist moral orientation that ascribes special moral significance to human beings on the ground that they are *natural persons*. It would not be easy to catalog them all. But even without doing so, I think it should be easy to see that the normative account of personal identity will invite us to re-think at least some of them. After all, it proposes several *metaphysical* revisions to our ordinary thinking about personhood and personal identity, and it stands to reason that some *moral* revisions will follow. As my title indicates, I will focus my efforts in this chapter on tracing some of the implications to do with personal responsibility.

It will be helpful to begin by acknowledging two sources of doubt about whether the normative account's specific metaphysical claims about personal identity – namely, that there can be multiple and group persons – are bound to have revisionist moral implications. First, it might be pointed out that, even if we grant that the possibilities of multiple and group personhood are in some sense *real* practical possibilities, they are not *available* to be pursued in the here and now unless they are *recognised*, and they *cannot* be recognized by anyone who doesn't believe the normative account from which they follow. But in the here and now hardly anyone *does* believe the account, which is precisely why it qualifies as revisionist! In the main, the persons who actually exist conceive themselves to be natural persons; and in the main they live accordingly – so it is unclear why the individualist moral orientation that goes together with that conception shouldn't be an adequate guide for the purposes of their actual moral dealings with one another.

Furthermore – and this is the second source of doubt that I want to acknowledge – suppose that the normative account does come to be believed, and that the possibilities of group and multiple personhoods do come to be realized. It might still be doubted whether this would require any *substantive* revisions of the individualist moral orientation, so much as a broadened understanding of the *field* of its application – so that it would apply to all individual agents regardless of their size. Arguably, this development would parallel what we have observed in the history of liberal

rights, whose field of application has been broadened time and again in response to suffrage and civil rights movements. Let me now address these doubts about the need for at least some moral revision.

The first doubt presumes that there are no *real* cases of group and multiple personhood for us to contend with in actual moral practice. But recall that, at the close of the last section, I suggested that the most *likely* case of group personhood would be a *hybrid* case that has a feature in common with multiple personhood – meaning that there would be a kind of segregation, or fragmentation, of rational efforts at the site of a single human life, so that *some* thinkings and doings help to constitute a group agent while others serve to constitute a smaller agent within that human life. There are many indications that we have *already* realized this hybrid case in actual fact – though it is easy for us to overlook, owing to our commonsense belief that each human being is a *natural person.* The belief naturally leads us to *misdescribe* any hybrid case that does arise as a *social* case of group agency in which each human 'member' of the group remains intact as an individual person whose personal identity, and sphere of personal responsibility, are unaltered by their involvement in the life of the group agent. The fact that they *are* misdescriptions shows up in the form of confusions in our practices, concerning how and where to assign personal responsibility for what.

One group endeavor in which I think this is evident is scientific inquiry. Any contemporary scientific investigation generally involves a whole group of human beings, whose thoughts and actions are all coordinated for the sake of a highly unified investigative project. Insofar as we recognize this feature of scientific investigation, we are prone to describe it as a *social* feature – that is, we are prone to describe scientific investigation as *teamwork*. But social relations arise among distinct persons when they engage one another from their own *separate* first-person points of view. And if we look closely at the exchanges that commonly take place in the context of a scientific investigation, we can see that questions are not being considered from separate points of view; they are being considered, rather, in the light of a whole body of pooled deliberative considerations. Some of these pooled considerations are given by a background theory, along with previous results achieved within the framework of that theory; some concern questions that the background theory leaves open; and some

concern what the current, ongoing, investigation has brought to light so far. These pooled considerations then provide a basis for deliberations, whose aim it is to work out their all-things-considered significance. I think that many scientists would agree that their exchanges should have the aim of working out the all-things-considered significance of such pooled considerations. But I don't think they necessarily have in view that, to realize this aim is to realize a condition that defines what it is for an *individual agent* to be fully or ideally rational. Correlatively, insofar as they do not appreciate this, they may fail to see that the pooled considerations from which scientific deliberations proceed constitute the single point of view from which a single investigating scientific agent reasons and acts *as one* – much as I described earlier in connection with the Philosophy Department. Yet at the same time, scientists *are* very well aware that contributions to science are not generally made by individual human beings, but rather by investigating *groups.* This fact is clearly reflected in their publishing practices, in which scientific papers are generally *co-authored.* Unfortunately, though, the fact does not mesh well with certain other academic practices that presuppose an *individualist* orientation. Consider, for example, how our universities award PhD's *only* to individual human beings. The standing rule is that the PhD. should be awarded to an individual *only if* they have *personally* made an original contribution to their field. Unfortunately, it is not generally feasible to separate out what any one human 'member' of a scientific investigative group is solely and uniquely *responsible* for having contributed. And even in cases where we *can* identify such a specific 'smaller' contribution, it does not typically meet the standards for the PhD. when considered by itself – only the larger contributions made by the whole investigating group agent do. So, in order that the PhD. can be duly awarded to an individual human being, a fairly artificial, and I would say *false*, scholarly attribution has to be made specifically to them, in spite of the fact that the result was *really* achieved by the entire investigating group agent. *And everyone in science knows this.* At the same time, scientists may be prevented by their standing beliefs from *describing* the situation in a way that fully accords with what they know. This would require them to take in what the normative account of personal identity reveals, which is that, typically, there *aren't* any individuals who are personally responsible for the outcomes of a scientific investigation

except the *individual group agent* from whose point of view the investigation proceeds – a group agent whose point of view and very existence are constituted for the sake of what it can achieve by reasoning and acting *as one.*

I can well imagine that many social ontologists would think they can readily assent to everything I just said, without giving up their view that human beings are natural persons, and that group agency is therefore always, inherently, *social* – in other words, that group endeavors never induce the sort of rational fragmentation that I described at the end of the last section in connection with the *hybrid* case of group agency. In response, I can only say, *but they do* induce such fragmentation, and so, we better adjust our moral thinking accordingly.

Elsewhere, I have discussed the Manhattan Project as a case in which the ideal of individual rationality *was* self-consciously pursued by a scientific group, owing to the direction of Oppenheimer. Whenever his different sub-groups set to *arguing* about why only *their* line was worth pursuing, he would generally re-direct their attention to the task of taking into account *all* of the results, from *all* of the work that had been done so far in the project, in order to work out what it would be best to do in the light of them *all.* If we look carefully at the Manhattan Project, we can see that it was, truly, a hybrid case of group agency – though as usual, the conception of human beings as natural persons can get in the way of appreciating the point. It is well known that many participants in the Manhattan Project began to have scruples about the project, but only one – Rotblat – actually withdrew. So perhaps it would be fair to regard him as living up to the ideal of overall rational unity within a whole human life. But in the biological life spans of the *other* human beings involved in the project, we can clearly see a segregation of rational effort. Some intentional activities were devoted entirely and exclusively to, and thereby helped to *constitute*, the single group agent that was working on the bomb, while other intentional activities constituted the life of another agent, smaller than human size. And it is quite vivid that some of the relations among intentional activities located at the site of a given human ‘member’ of the project were genuinely *inter*-personal – insofar as there was active working on a bomb that proceeded from the point of view of the larger group agent, which occurred alongside active *disapproval* of that very work from the quite different point

of view of another agent, smaller than human size, who was a husband and father and citizen.

It might be objected that it would be morally *better* to insist upon viewing all of the human participants in a project like the Manhattan Project as *natural persons*, and to regard each as a locus of responsibility for all that they individually did, where this would be understood to include their contributions to the project. As I noted earlier, List and Pettit believe, or anyway *hope*, that this individualist orientation will help to secure morally better outcomes – the hope being that any human being with enough conscience and integrity would refuse to participate in such a project. But it is obvious that *withdrawing* from the project would not suffice to *stop* the project – that would require a much bigger effort that would address the point of view from which the project was undertaken, not to mention *funded*. So, my response to this objection is that it would be *even better*, morally speaking, to engage in such a much bigger effort.

So far, I have been emphasizing that, if we think – as I do – that holding others responsible involves adopting Strawson's participant stance, then we need to attend more carefully to the task of correctly identifying the individuals whom we are holding responsible. We cannot merely assume that every time we are faced with an individual human being, we are presented with an individual agent of human size whom we can engage *as one*. In our efforts to engage with others, we must learn to track where the commitment to reasoning and acting *as one* is really falling with respect to human being – so that we can properly identify the *points of view* that we are trying to engage.

I have just been discussing one clear *revision* of our moral practices that is called for normative account of personal identity. But there are *other* revisions in the offing, that have less to do with identity *per se* – less to do with the possibilities of group and multiple personhood – and more to do with other claims of the normative account, most especially: the reductionist claim that the life of a person consists in *nothing but* the thinkings and doings that constitute the life of the person; and the corollary claim that agents are *self-constituting*. In combination, these claims entail that the ultimate loci of agency itself are the particular thinkings and doings that together constitute the life of an agent. These claims put real pressure against a very longstanding practice in the western tradition, of holding

individual agents responsible for their *past* actions, for the purposes of both *judgment* and *sanction* (reward or punishment). As Locke notes, this requires that we make sense of what it is for the *same* person to have performed those earlier actions and then be 'held to account' for them at a later time. While *he* claims that all that is required is *sameness of consciousness*, the normative account requires a *normative* relation of *ownership*, which is embedded in a single *rational* point of view. When the account goes on to portray the agents who possess such a rational point of view as *self-constituting*, this seems to afford them an easy *escape* from eventual judgment and sanction for their past actions. It seems that all they would need to do is to *end their existence* – not by biological death, but rather by redrawing the boundaries within which future thoughts and actions are to proceed. Once such new boundaries are forged, new lives can go on, while the agents who had performed the past actions will no longer be available for judgment and sanction. A similar complaint is often raised about corporations, construed as a social case of group agency: they *clearly* have such an escape route, in the form of *disbanding* the corporation. This worry about corporations fuels a wish to find an ultimate locus of personal responsibility – an individual agent who will still exist, and who can still be held to account, even after the corporation has disbanded. And then the conception of the human being as a *natural person* gets wheeled in to fill the bill. A similar worry is often raised about the case of multiple personhood. If it is possible for multiple persons to constitute themselves within a single human life, won't this afford human individuals with a ready escape route from personal responsibility? This worry fuels the same wish as before, to find an ultimate locus of personal responsibility, and once again, the conception of the human being as a *natural person* gets wheeled in to satisfy the wish – because in the case of such natural persons, the biological boundaries that circumscribe their lives, and therewith, their own spheres of personal responsibility, cannot be renegotiated. To many, this seems *crucial* to any satisfactory moral practice of judgment of, and punishment for, past actions.

I must confess that, until recently, I did not concern myself very much with these sources of resistance to the normative account of personal identity. As far as I was concerned, it would suffice to point out that even *natural persons* die – so unless there is life after death and a divine

judgment, we shall have to cope with the obvious fact that some agents do pass away before they can be held to account, and all that the normative account of personal identity adds is a *new way* to pass away! But after reflecting more fully and soberly on the issue, I can see that the normative account is bound to disturb aspects of our longstanding practices of holding individual persons responsible. This comes into especially interesting focus when we consider that we adopt Strawson's participant stance from our own first-person standpoint as agents. We do not observe, predict, or offer causal explanations from this standpoint. Rather, we apprehend occasions for agential response *on our part,* in response to matters of value. When we are in agential mode, we recognize that our agency cannot reach into the past. This point holds even when we adopt the participant stance towards others, in order to evaluate *their past* actions. Insofar as we aim to *engage* another agent via our reactive attitudes, we are looking for *uptake* from their end, which would have to take the form of an agential response on their part *in the present.*

These aspects of personal agency may seem to invite a consequentialist moral orientation that is more focused on the changes that can be brought about in the here and now than on moral judgment and sanction with regard to past actions – unless, of course, such judgment and sanction may have good consequences going forward. If this were true, then some of the moral revisions called for by the normative account of personal identity would be akin to ones that Parfit (1984) calls for in connection with his psychological reductionist account, when he argues that *identity is not what matters,* and that what really matters is the greatest aggregate well-being, impersonally conceived without regard for *whose* well-being might be at issue. But this is *not* the lesson of the normative account. It does *not* imply that agency ought generally to be directed at securing the greatest aggregate well-being, or for that matter, at securing well-being at all. It says, more neutrally, that individual agents constitute themselves for the sake of the *ends* that their existence makes it possible to pursue – which I call a *unifying project,* commitment to which calls for reasoning and acting *as one* within some specific boundary that is not biologically given with a human life but is set by the nature of the project. It so happens that there are all sorts of worthy projects that may come into view in different times and places, depending on the different opportunities afforded in them - after all, differently

situated agents would be able to cultivate and use different sorts of capacities, and to produce different kinds of artifacts and outcomes, and to aspire to different kinds of knowledge, etc. I don't see any reason to adopt the consequentialists' narrow, largely *hedonic*, conception of the value of these projects, on which their ultimate purpose must be to produce well-being; nor do I see any reason to privilege the consequentialist goal of contributing to the greatest aggregate well-being over all other goals.

Nevertheless, I do think that the normative account of personal identity does imply that the best way to think of personal responsibility is forward-looking. To exercise agency *at all* is to *take* responsibility for some small part of what will happen in the world. Human lives *can* be the sites of impulsive behavior of the sort that Harry Frankfurt (1971) labels *wanton***.** Or, they *can* be sites where personal agency comes to be exercised in a more reflective and coordinated and normatively driven way. According to the normative account, this is how individual agents come to constitute themselves, and in doing so, come to forge first person points of view from which to reason and act *as one*. I think it is obvious that the *first* such efforts, in the time of youth, would produce agents whose lives span much less than a whole human life. I find it equally plausible that the first such youthful efforts would produce agents whose lives span across human lives. For I simply do not share the prejudice that the natural way for personal agency to be exercised is for the sake of ends (unifying projects) that would confine it within *separate* human lives. I think the sense that this is natural comes from a history in the West, in which the importance of the *human individual* took on a succession of meanings – a creature to be judged by God, a site of personal conscience and integrity, a locus of individual rights and liberties, a self-interested agent in a marketplace, a romantic project of self-fashioning (I'm sure I'm overlooking further meanings!). But to return to the point about personal responsibility being *forward looking:* If we really are convinced, as I myself am, that Strawson is right to say that responsibility is something that can *only* be apprehended from the vantage point of the participant stance, which is the *engaged* stance of an *agent*, then it makes sense to retain a forward-looking interest in it. And when we retain a forward-looking interest, the fact that the boundaries between individuals are not fixed, but can always be renegotiated, appears less as a threat to our familiar practices of praise, blame and punishment, and

emerges more positively as an invitation to entertain the whole range of practical possibilities before us.

A final note in closing: If we wish to preserve a more retrospective, and even retributive, focus when we attribute personal responsibility to others, then the lesson of this chapter is that we must not forget the following: just as Strawson argues that the agency and responsibility of others is something that can be apprehended only from the vantage point of the participant stance, I am arguing that the same holds for the very *identities* of other agents. The boundaries of their lives *as agents* are not given with human biology and phenomenology, but are forged for purposes; and it is only from the vantage point of the participant stance, and through attempts at engagement, that we can discern where those boundaries are, and from *which* points of view thought and action proceed. If we bear this in mind, we will find that *many* cases of group agency are *not* social. And if we bear this *duly* in mind, we will cease to lapse into the problems and confusions and incoherences I registered above, about where to locate personal responsibility.

References

Bratman, M. 2014. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.⏎

Frankfurt, Harry G. 1971. ‘Freedom of the Will and the Concept of a Person’. *The Journal of Philosophy*, 68(1), 5–20.⏎

Gilbert, M. 1989. *On Social Facts*. Princeton: Princeton University Press.⏎

Hume, David. 1888. *A Treatise of Human Nature*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press.⏎

List, C. and Pettit, P. 2011. *Group Agents: The Possibility, Design and Status of Corporate Agents*. Oxford: Oxford University Press.⏎

Locke, John. 1975. *Essay Concerning Human Understanding*, ed. P. Nidditch. Oxford: Oxford University Press.⏎

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.⏎

Rovane, C. 1998. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.⏎

Rovane, C. 2014. 'Group Agency and Individualism', *Erkenntniss*, 79: 1663–1684.⏎

Rovane, C. 2025. 'Freedom and Unfreedom in Human Categories: The Case of Multiplicity', in Akeel Bilgrami and Jonathan R. Cole (eds), *Non-Coercive Threats to Academic, Political and Economic Freedom*. Columbia: Columbia University Press.⏎

Strawson, P. 1963. 'Freedom and Resentment', *Proceedings of the British Academy*, 48: 187–211.⏎

Strawson, P. 1985. *Skepticism and Naturalism: Some* Varieties. Columbia: Columbia University Press.⏎

OceanofPDF.com

5 Collective Obligations Without the ‘Quantum of Blame’

Bill Wringe

DOI: 10.4324/9781003612827-6

1 Introduction

Can there be (forward-looking) obligations falling on non-agential groups? Many fınd this suggestion implausible. Nevertheless, there is a natural strategy for arguing in favour of such obligations: namely to argue that in some cases, an obligation must fall *somewhere* and then give reasons for thinking that the obligation cannot fall on any particular agent, either individual or non-individual. Elsewhere I have argued individual subsistence rights give rise to obligations falling, collectively, on humanity as a whole (Wringe 2005; Wringe 2014; Wringe 2018). The argument depends on the idea that rights give rise to correlative obligations, and the claim that humanity as a whole is the only plausible obligation bearer here. Christine Korsgaard has recently advocated a similar strategy (Korsgaard 2018).

One important and much-discussed objection to this line of argument depends on the idea that a correct understanding of the principle ‘ought implies can’ rules out the possibility of non-agential obligation bearers (Pinkert 2014; Lawford-Smith 2015; Collins 2019; Collins 2023; Schwenkenbecher 2021a). I have argued that it does not (Wringe 2020). Here I want to discuss a different kind of objection to it. This involves examining parallels between that strategy and a way of arguing for the existence of backward-looking responsibility, and identifying a flaw in the backward-looking strategies which suggest a parallel weakness in the corresponding argument for forward-looking collective obligations.[1]

The suggestion is that the argument in the backward-looking case involves making a mistake about the attribution of blameworthiness: it assumes that in certain circumstances we can establish that blame is to be attributed independently and prior to establishing where that blame should fall. Call this the Quantum of Blame error.[2] There is room for debate about whether the Quantum of Blame error really is an error, and whether it is being committed in any particular case. However, my primary interest here is to argue that whatever one thinks about this, what I shall call the Wringe-Korsgaard argument for forward-looking responsibilities is not vulnerable in the way that this line of thought would suggest. The argument will turn on distinguishing between the functions of blame and those of accountability. The moral is that arguments from the existence of accountability gaps are more powerful than arguments from responsibility gaps.

2 The Rights-Based Argument for Collective Obligation

I start by setting out as briefly as possible the argument for non-distributive collective obligations that plays a central role here. It starts from considerations about rights, and from a version of what is often called the 'correlativity thesis'. According to the version of the correlativity thesis I am interested in, an individual's rights generate obligations on other people to see that the rights in question are respected (O'Neill 1986; Collins 2016). In the case of what are sometimes called 'negative' or 'liberty' rights, the obligations generated include an obligation on each individual – indeed, on each moral agent, if one thinks as I do that there are moral agents that are not individuals – to act in ways that do not transgress that right. Thus, for example, my having a right to life entails, among other things, an obligation on others not to kill me. It may also include an obligation on others not to expose me to unreasonable risks of death. It may also, more controversially, entail a suite of other obligations, including obligations to impede others from interfering with my right to life, to punish those who transgress, to uphold effective institutions and so on.

The correlativity thesis is widely accepted, but not entirely uncontroversial (O'Neill 1986; Sen 2004; Collins 2016). Importantly, though, even weak versions of the correlativity thesis are sometimes thought to raise difficulties for the idea of 'positive' or 'welfare' rights, such as a right to subsistence or to healthcare (O'Neill 1986; Srinivasan 2012). For if the correlativity thesis is correct, the obligations it gives rise to must have bearers. Where liberty rights are concerned, specifying who these obligation bearers are is relatively straightforward. The obligations in question fall on *each* moral agent.

Matters are more complicated when we consider positive or welfare rights. The idea that a right to an acceptable basic standard of living generates a duty on *each* individual to see to it that they have food seems implausible. Nor is it clear that we can find a non-arbitrary way of specifying a set of individuals on whom the duty in question might fall. In fact, according to some opponents of the existence of welfare rights, we cannot find *any* acceptable way of specifying the obligation-bearers. But since there cannot be obligations without obligation-bearers, the correlativity thesis seems to entail that there cannot be such obligations.

However, this is too quick. The argument overlooks the possibility of obligations falling somewhere other than one or more individual (or collective) agents. The neglected alternative is that the obligation in question falls on everyone: not, as in the case of obligations arising out of liberty rights, on everyone *distributively*, so that each individual has the obligation in question, but in a *non-distributive* manner. This non-distributive collective obligation might then be taken to ground a range of obligations on individuals, thereby escaping the charge that it fails to be action-guiding, and therefore, arguably, fails to be an obligation at all.

One might have various qualms about this argument. One might, like Amartya Sen, find the correlativity thesis implausible (Sen 2004). One might be dubious of the (supposed) distinction between positive and negative rights. Alternatively, one might object, as Thomas Pogge has, to the suggestion that some kinds of welfare right are correctly characterised as negative rights (Pogge 1992). One might also worry about the conclusion of the argument. Is there a plausible and non-arbitrary way of specifying exactly who counts as falling under the heading 'all of us'? Can obligations fall on non-agential groups (Lawford-Smith 2015), or on large groups

which have no shared goals (Collins 2019) or obvious means of co-ordination (Collins 2019; Schwenkenbecher 2021a; Schwenkenbecher 2021b; Schwenkenbecher 2022)?

Furthermore, even if one accepts that obligations *can* fall on non-agential groups, one might have concerns about this particular (putative) obligation-bearer: one might think that the things that it is (supposedly) obliged to do are things that it cannot (in the relevant sense of cannot) do (Lawford-Smith 2015). One might also worry that the conclusion of the argument is incompatible with a stronger version of the correlativity thesis, on which the obligations generated by rights must be claimable: that is to say, must identify one or more agents who can be compelled to act; (Collins 2016) and so on.

While I think that each of these objections can be countered effectively, I shall not discuss them here. Instead, I shall focus on ways in which this line of argument runs parallel to a kind of argument which has often been used to defend the existence of non-distributive backward-looking responsibility: so-called 'Responsibility Gap' arguments. Arguments of this sort are sometimes criticised for making a tacit (and implausible) presupposition, we can establish that someone must be to blame for a bad situation, without knowing who. If this is a flaw in Responsibility Gap arguments, we might worry that parallels between such arguments and the Wringe-Korsgaard argument for non-distributive forward looking obligations give us reason to suspect a similar flaw in the latter. I shall argue that they do not.

3 Responsibility Gap Arguments

Many arguments for non-distributive collective responsibility seem to involve the following three-step strategy:

STEP 1: Describe a situation in which it seems intuitively plausible that some or all of a group of individuals are responsible for a bad state of affairs.

STEP 2: Argue that none of the individuals involved in that outcome can appropriately be held responsible for that state of affairs.

STEP 3: Identify some further thing to which the individuals in STEP 1 are appropriately related and argue that the responsibility which we find intuitively plausible to assign to them falls on this further entity.

In some versions of the argument the further thing to which responsibility is assigned is a collective agent whose existence can be recognised independently of the assignment of responsibility which takes place in STEP 1 (Pettit 2007). In other versions the assignment of responsibility to a further entity provides us with reasons for recognising that entity as a moral agent (Copp 2006). And in others, the entity to which responsibility is assigned is taken not to be an agent, and the argument is taken to establish the possibility of assigning responsibility to non-agential groups (Chant 2015).

Arguments of this sort are known as 'Responsibility Gap' arguments. They have been widely criticised (Ludwig 2007; Miller 2007; Hindriks 2009; see also Copp 2012). One of their weaknesses is that they presuppose that we can identify an outcome as one for which some entity or other bears responsibility without establishing which entity that is. This is at odds with central aspects of our practice of assigning responsibility: in particular, with the fact that judgments about the ways in which agency is implicated in those outcomes can lead us to withdraw attributions of responsibility altogether. The plausibility of these arguments depends – I shall claim – on a natural, but nonetheless clearly mistaken thought about the nature of blame, which I call the 'Quantum of Blame' error.

The Wringe-Korsgaard has the same kind of structure as Responsibility Gap arguments. We might describe it as appealing to the possibility of 'accountability gaps'. If so, we might think it involves an error analogous to the 'Quantum of Blame error'. At the very least, those who are persuaded that the Wringe-Korsgaard argument is plausible need to explain why we should treat the two kinds of argument differently.

I shall argue that there is an explanation. It is based on differences between blame and accountability. The 'Quantum of Blame' error is, I shall argue, a mistake about blame – or about the kinds of practices which would best fulfil those useful functions which the notion of blame performs for us. Forward-looking notions of obligation are less closely tied to the notion of blame than the backward-looking notion of responsibility which figure in

Responsibility Gap arguments. They are more closely tied to a forward-looking notion of accountability. Furthermore there is a principle involving the forward-looking notion analogous to a (mistaken) principle implicated in the Quantum of Blame error, but which unlike that principle is highly plausible.

4 Some Paradigmatic Arguments

Here are some apparent examples of the three-step strategy. Here, first is Philip Pettit discussing the 'Herald of Free Enterprise' disaster. In arguing that it may make sense to hold the ferry operator legally, as well as morally responsible, Pettit writes:

> I hold that even when all the relevant enactors in a group action have been identified and held responsible, still it may be important to hold the group agent responsible as well…. The responsibility of enactors may leave a deficit in the accounting books, and the only possible way to guard against this may be to allow for the corporate responsibility of the group in the name of which they act.
>
> (Pettit 2007: 194)

Here is David Copp, talking about what he calls the 'Collective Moral Autonomy Thesis'.

> (BOARD) Suppose that Bob, Carol, Ted, and Alice are the members of the Board of Governors of the state prison. A month ago, the Board voted against strengthening security procedures at the prison, and it is now being blamed for this decision in light of bloodshed resulting from a recent attempted escape….(1) It would be unreasonable to blame Carol since she brought forward the motion to increase the funding, she argued for the motion, and she voted for it. Bob

missed the meeting. He had a very good excuse, since he was out of town dealing with a family emergency. Alice voted against the motion, but she had an excuse for doing so. She and Ted agree on most issues, and there are too many difficult issues that come before the Board for both of them to do thorough research on every issue. They have a standing agreement to take turns doing the necessary research, and it was Ted's turn to do the research. Alice voted against the motion, seeing that Ted was also voting against it, and believing that Ted had done the research needed to make a reasonable decision…Suppose we learn that Ted had an excuse…Suppose that Ted has been under enormous stress because of serious illness in his family and as a result he has been unable to sleep. Given his exhaustion, he mis-remembered whose turn it was. He believed it was Alice's turn. When the time came to vote, he was confused as a result of his exhaustion, and he thought that Alice was signaling him to vote against the motion… In this case, it seems to me that Ted is not blameworthy for voting against the motion. (2) It also seems, however, that it would be a mistake to absolve the Board from blame for the decision (1)….I think that the Board is appropriately blamed, all things considered.

(Copp 2007: 377–379)

Finally here is Sara Chant arguing for the possibility of moral responsibility falling on non-agential groups (Chant 2015):

(STAND-OFF) Mr Good, Mr Bad and Mr Ugly find themselves in a 'Hollywood Standoff'. Each is holding a gun to the head of another member of the group such that everyone in the group has a gun pointed to his head. Everyone knows

> that the first to shoot will surely be shot, and everyone believes that everyone knows this. Moreover, everyone believes that if anyone were to begin to lower his gun, chaos could ensue and everyone could get shot. Additionally, nearby, and within clear sight of the men, a child is suffering from severe dehydration. The men see that the child is in imminent danger and could be easily saved if they all lowered their guns and gave the boy the last few sips of water from each of their flasks. Nevertheless, Mr Good, Mr Bad and Mr Ugly uphold the standoff and the child dies…
>
> ….according to the case, the men together could have put down their guns and given the child the last few sips of water from each of their flasks. So even though each has an excuse that absolves him of distributive responsibility, those excuses do not thereby absolve them jointly. Thus, The Hollywood Standoff is a prima facie example *of non-distributive responsibility without a collective agent*.
>
> (Chant 2015: 86)

5 The ‘Quantum of Blame’ Error

We can easily imagine situations analogous to those presented in the arguments above in which we might be initially inclined to assign responsibility but in which further scrutiny of the circumstances established that no entity – that is to say, no individual and no further entity either – should be held responsible. Suppose that investigation established that the sinking of the Herald of Free Enterprise had been caused not (as in reality) by someone having left open the car deck doors, but by a freakish (and unpredictable) natural event in which the waters of the Thames suddenly became strongly acidic and corroded the properly closed doors (HERALD*); or that the Board of Governors votes electronically, and that,

because of an electronic glitch, their vote to increase staffing levels is misrecorded (BOARD*). Or suppose, again, that the death of the child in the Hollywood stand-off is pre-emptively caused by a rock fall caused, not by the actions of Good, Bad and Ugly but by a freak gust of wind (STAND-OFF*).

In HERALD*, BOARD* or STANDOFF*, it is easy to resist the thought that responsibility for a bad outcome needs to be allocated somewhere. Our initial judgement that there is some responsibility to assign is undermined when we discover the full facts of the situation. Talk of a ‘gap in the accounting books’ seems out of place here: if the individuals involved in these scenarios are not appropriately held responsible, there is no need for us to find some alternative locus of responsibility.

There might still be reasons to find individuals and/or groups responsible for *something* in these situations. However, we will not be allocating responsibility for the same thing as we find natural in HERALD, BOARD or STAND-OFF. We may also be inclined to attribute some kind of responsibility to make good any harms suffered as a result of HERALD*, BOARD*, and STAND-OFF*. But if we do so, it will not be as a result of any individual or any group’s backward-looking responsibility for the kind of state of affairs for which we seek to find a responsibility bearer in HERALD, BOARD, or STAND-OFF.

Someone who insisted that in these cases this backward-looking responsibility must be assigned somewhere would be making a mistake. The mistake is that of thinking we could assign responsibility for what happens in these scenarios in a way that was independent of, and prior to being in a position to assign that responsibility to any particular bearer. There is a mistake one might make about blame which runs parallel to this mistake about responsibility. It is to think that in a situation of this sort, someone (either singular or plural) must be to blame.

On some conceptions of the relationship between judgments of moral responsibility and judgments of blameworthiness these two mistakes are the same. Furthermore, one way of defending the idea that there is no need to assign moral responsibility in cases like this derives from the idea that in a situation like this no one would be to blame. So I shall call the error committed in thinking that there must be some responsibility to be assigned in these cases the ‘Quantum of Blame’ Error.

The starred scenarios differ from the originals on which they are variants in an important respect. Whereas the first set involve situations in which agency is implicated, the second involve situations which following Caroline Arruda, one might call 'mere events' (Arruda 2017). So we might think that the mistake being made when someone assigns responsibility in these cases is that of assigning responsibility for mere events, rather than for actions.

However, this is not the full story. Assignments of responsibility can also go awry in a way that does not involve mistaking mere events for situations in which agency is implicated. Suppose that in the Herald of Free Enterprise case the failure to shut the doors of the ferry can be traced to a single individual who has a justification for attending to something else: at the moment when the doors needed to be shut they were engaged in preventing some other – unpredictable but imminent – disaster (Call this HERALD**). Or suppose that the electronic voting machines had been accidentally miswired in such a way that the results of their vote was the reverse of what they should have been (BOARD**). Or finally, suppose that the inaction of Good, Bad and Ugly in Hollywood Standoff is excused not by a threat, but by the fact that if any of them makes a sudden move they risk setting off an avalanche that will kill all four of those involved in the predicament (STAND-OFF**).[3]

Actions are implicated in each of these scenarios. Individuals can be called upon to explain themselves in these situations, and the demand is satisfied, in each case by the agents being able to provide an excuse for their action or inaction. If these scenarios involved only what Arruda calls 'mere events', there would be nothing to give an account of, and no-one to whom the demand might appropriately be addressed. Still, it would be just as inappropriate to assign moral responsibility to any entity other than the individuals involved in HERALD**, BOARD** and STANDOFF** as it would be to do so in HERALD*, BOARD* and STANDOFF*.

Just as in the starred scenarios, it may be that in each of the double starred scenarios we might want to try to find some agent or group responsible for *something*. Did someone assign crew members that meant that one individual was overloaded with responsibilities? Had a maintenance crew skimped on a scheduled inspection of the voting

machines? Had Good, Bad and Ugly recklessly ventured into an area prone to avalanches? These are all coherent ways of developing the scenarios. But they are not the only possibilities, and we should be wary of assuming in every such development blame for a tragic outcome must be attributable to some agent.[4]

Perhaps the situation that led to the sinking involved a situation which no-one could have been expected to anticipate, such as a sudden outbreak of a debilitating illness among the crew. Perhaps the maintenance crew reasonably believed that this set of machines had already been checked – perhaps it had indeed been checked and only malfunctioned subsequently. Perhaps Good, Bad and Ugly are not three antagonists, but three rescuers sent from different towns in search of a missing individual, each believing that they need to protect themselves from potential assault. And so on. In each case we have a situation where individual actions are implicated but where it turns out our initial impulse to place blame is undermined by further knowledge of exactly how the tragic situation came about.

6 Does the Three Step Argument Involve a Quantum of Blame Error?

Anyone who sought to find a locus of blame in either HERALD*, BOARD* and STAND-OFF* or in appropriate developments of HERALD**, BOARD** and STANDOFF** would be committing an error. The three stage argument which I formulated in Section 2 involves a mistake of the same sort. For we could formulate versions of the three-stage argument which would cover the starred examples. However, one might still doubt whether the arguments cited in Section 3 are most charitably understood as instantiating that form. Let us examine them more closely. I shall argue that in each case, the Quantum of Blame picture seems to be present in the background.

6.1 Pettit

One reason for taking the Quantum of Blame idea to play a role in Pettit's thinking derives from his use of the metaphor of the accounting book. Pettit suggests that in a case like HERALD, we should find something other than

an individual agent to which we can attribute responsibility in order to avoid ‘a deficit in the accounting book’. When we settle accounts we are trying to allocate a quantity of money – something which exists independently of any particular bearer (though not, of course, to the general social background against which the institution of money makes sense) – to its rightful owner. The metaphor suggests a commitment to the idea that we can figure out how much responsibility there is to be allocated prior to, and independently of any attribution of that responsibility to a particular bearer. To the extent that the metaphor of the accounting book plays any persuasive role in Pettit’s argument, the quantum of blame seems to be in the background.

However, Pettit’s position is more nuanced than this. In arguing that judgments of corporate responsibility are appropriate in the *Herald of Free Enterprise* case, Pettit writes:

> The reason for this, very simply, is that it is possible to have a situation in which there is ground for holding the *group agent* responsible, given that it satisfies the three conditions listed, but not the same ground for holding individual enactors responsible.
>
> (Pettit 2007: 196)

This suggests that in situations where no group agent was present, it would be inappropriate to assign responsibility to some entity other than individuals involved in the situation. Several of Pettit’s comments about non-agential groups point in the same direction. For example he writes as follows:

> Shortfalls of individual responsibility have a distressing aspect in the case of the unincorporated collection, since they mean that although the individuals do something bad together, there is no-one to hold responsible.
>
> (Pettit 2007: 196)

> As individuals, the members of an unincorporated collection are agents, but in cases of these kinds they are not fully culpable agents,…. For the collection that they constitute, being unincorporated, is not an agent. It fails the first condition for fitness to be held responsible.
>
> (Pettit 2007: 195)

In short, when the conditions for a groups being fit to be held responsible are not fulfilled, Pettit would concede that there is no locus of responsibility. These conditions are far from trivial: Pettit suggests that in order to be fit to be held responsible a group must be an autonomous agent, with access to evidence required for making choices between values, and control over its actions of a sort that enables it to make choices about how to act on the basis of judgments about value (Pettit 2007: 177ff). Many groups will fail to meet these conditions, as Pettit notes.

But admitting this need not be taken as undermining a commitment to the accounting book metaphor. Pettit encourages us to see these cases as ones where the gap in the accounting books simply cannot be filled. Furthermore, Pettit describes the shortfall of responsibility in the case of unincorporated groups as 'distressing' (Pettit 2007: 196). There is, in his view, something amiss with such situations. Without the 'Quantum of Blame' picture in the background, it is difficult to see why we should regard this aspect of the situation as 'distressing': if there is no responsibility to allocate prior to our identification of bearers of responsibility, there is no reason to be distressed by our failure to find anyone to allocate it to. (Of course, there are other things to be distressed about here: most obviously, the massive loss of life involved. But what Pettit describes as being distressing is, in particular, the shortfall in *responsibility*.)

Furthermore this aspect of Pettit's position is not merely a rhetorical flourish: on his view, the fact that these shortfalls in responsibility are distressing provides us with a reason to treat some non-agential groups as though they were responsible in what Pettit describes as a 'proleptic' fashion. The view seems to be one on which we are encouraged to bring responsible agents into existence to avoid shortfalls. But without the

Quantum of Blame picture in place, it is unclear why we should see the absence of a blameworthy individual as involving a 'shortfall' at all. In short despite the apparent nuance in his position, Pettit does seem committed to the Quantum of Blame picture.

6.2 Copp

Copp's argument seems to involve an example of the three-step strategy in the following way: in step 1, we identify the break out as a state of affairs for which blame is appropriate; then we establish that no individual is to blame for that outcome – the discussion of Alice, Bob, Carol, and Ted (step 2). We then conclude that some further entity – namely the Board – is responsible for the outcome (step 3). Since none of the members of the board is responsible, this is a case of *non-distributive* moral responsibility.

One might still take my attribution of the three step strategy to Copp to be uncharitable. First, Copp's argument is formulated in terms of blame, rather than responsibility. However, this is something of a red herring: careful reading of the passage reveals it to be concerned with questions about the circumstances under which we *should* blame various kinds of entities (he says for example, that it would be a *mistake* to absolve the Board from blame). And I take it that Copp takes these facts to be relevant to the claims with which his paper is concerned because the facts about whom we should blame correspond to or mirror facts about responsibility.

Furthermore, Copp's description of the situation as one in which the Board is blamed in the light of bloodshed resulting from decreased security makes it natural to read the situation as being one in which the Board is blamed not merely for its decision, but for the results of that decision: reduced staffing, bloodshed and so on. If these features of the situation aren't results of the Board's decision then, arguably, the Board is being scapegoated. So even if Copp's argument does not exemplify the three-step strategy, his example certainly suggests an argument that can be understood along these lines (and it is reasonable to suspect that at least some of those who have been persuaded by the argument have in fact understood it along the lines I am suggesting here).

Copp limits his discussion to cases where something that we might plausibly regard as a collective agent – namely, a committee – is in play. It

is far from clear what Copp would say about situations like the one which Chant considers, where there is, plausibly, no collective agent for the responsibility from which Bob, Carol, Ted and Alice are absolved to fall on. So one might be tempted to absolve Copp from any suspicion that he is implicated in 'Quantum of Blame' style thinking.

However, Copp argues that scenarios of the sort he describes as giving us reasons to recognise that entities such as committees are, in fact, agents (Copp 2006: 214ff). They must be agents, he suggests, precisely because, in cases such as BOARD, they are loci of non-distributive moral responsibility. This argumentative strategy suggests that we can recognise that the Board is responsible prior to our recognition of it as an agent: if we could not do so, the our attribution of responsibility could not provide us with grounds for thinking that the Board constitutes a collective agent. But to think this is, quite precisely, to be committed to the Quantum of Blame picture.

6.3 Chant

It seems clear that the passage from Chant exemplifies the three step strategy. The case she describes seems to have the same kind of structure as Copp's: a situation where we wish to assign responsibility for an outcome, but are blocked from assigning it to any particular individual. (We might wonder why Chant holds that Messrs Good, Bad and Ugly each have an excuse here. Her argument relies on what she calls the Threat Principle. This principle 'absolves an agent of moral responsibility when he acts solely because of serious threat in the same way that coercion absolves an agent of moral responsibility when he acts solely because of coercion' (Chant 2015: 3). Since 'each continues to threaten another man in the group only because he himself is under credible threat', she concludes that 'the Threat Principle absolves each man of moral responsibility for his role in maintaining the standoff.') Again, we see the three steps: identify a bad outcome, for which someone is to blame; establish that none of the individuals concerned can be blamed; place the blame on some entity other than an individual agent.

There is a significant disanalogy between Chant's example and Copp's. Whereas the former involves a group which is plausibly viewed as

constituting a collective agent (at least on standard conceptions of what collective agency involves), the latter is deliberately constructed in such a way as to involve a group which it is intuitively implausible to regard as involving a collective agent. One consequence of this disanalogy is that the argument which we considered above for thinking that Copp's argument does not exemplify the three-step strategy seems inapplicable here.

Since the collection of Good, Bad and Ugly does not constitute an agent, we cannot claim that the argument is plausibly construed as one on which the thing for which Good, Bad and Ugly are held responsible is an action in some sense where actions are contrasted with outcomes. If there is no agent, there is no action. Nor, I think, can we say that they are held responsible for an omission – at least in the sense in which omissions just as much as actions can be understood as the upshot of agential capacities. If Good, Bad and Ugly are excused for their individual omissions (as Chant holds), there is no further omission – in this sense – for which we can hold Good, Bad and Ugly collectively responsible. There is just the upshot – the outcome – of their individual omissions.

7 From Responsibility to Blame

The arguments considered in Sections 3–5 concern situations where we are trying to assign responsibility for something that has (already) gone wrong. They are, we might say, backward-looking. We might also describe the arguments in question as 'blame-based', since there seems to be a close relationship between the assignment of responsibility and the judgment that blame is appropriate (Strawson 1962; Wallace 1994; Shoemaker 2015; Shoemaker 2017).

Here are a range of different views about the relationship between the assignment of responsibility and the judgment that blame is appropriate. We might think that the judgement of responsibility grounds the appropriateness of blaming (Brink and Nelkin 2013). If so, the appropriateness of blaming something other than an individual provides us with a means whereby we can *recognise* the existence of collective responsibility. Alternatively, we might take the judgement of responsibility to be grounded in our recognition of the appropriateness of blame. On a

view like this, the appropriateness of blame is more fundamental than the judgement of responsibility. Or, finally we might take the fact that blame is appropriate and the fact that we assign responsibility in this case to be exactly the same fact (Shoemaker 2017). On this view, recognition of the appropriateness of blame is this situation just is recognition of moral responsibility.

Whichever view we take here, the connection between blame and moral responsibility in these cases helps us to see not just that the Quantum of Blame error is an error, but *why* it is an error. There is, of course, a great deal of dispute about the nature and purpose of blame (Coates and Tognazzini 2012, Tognazzini and Coates 2021). Here, I shall understand it to be an expressive and communicative practice that has among its purposes those of establishing moral responsibility, morally educating both its targets and those who are witnesses to its expression, and expressing solidarity with its victims. (Others who have emphasised the communicative and educative dimensions of blame include Duff 1986; McKenna 2013; Bennett 2013; MacNamara 2015; and Fricker 2016.) I leave open the question of whether it is the best, the only, or even an especially effective way of achieving these purposes. It is important, however, that those practices are somewhat malleable – so that we can assess how well they serve these goals, and how they might be adapted to serve them better and in ways that best serve other human purposes (McGeer 2013). Judgments about the appropriateness of blame will be subject to pressure exerted by recognition of ways in which the practices serve those practices or fail to serve them.

Seeing blame in this way helps to explain why we are, and should be, resistant to regarding anyone as responsible for bad outcomes in any of HERALD*, BOARD*, STAND-OFF*, HERALD**, BOARD** and STAND-OFF**. For it is plausible that there would be something dysfunctional about practices which assigned blame in cases like this. There is we might think, no need to morally educate anyone in scenarios like this, nor to stand in solidarity with the victims in the way that we stand in solidarity with the victims of wrongdoing. Nothing is gained, and much is perhaps to be lost, by a gesture expressing willingness to defend them against the further deeds of the agents implicated in these situations. (Other kinds of solidarity, such as those expressed by regret and remorse may

nevertheless be appropriate.) If that is right, then there is nothing arbitrary about our withholding judgments of responsibility in situations like this.

Matters are different in HERALD, BOARD and STAND-OFF. In these situations, their respective authors suggest, there is an appropriate target of blame. I have no desire to question that judgment, and no reason to disagree that the situations involved are ones where a group is non-distributively responsible. However, an appreciation of the connection between responsibility and appropriate blame helps us to see why arguments which fit the three-step pattern, are of no use in helping us to establish this, and, relatedly, why Pettit's metaphor of a gap in the accounting book is inapposite. For it is plausible that any set of practices which started from the assumption that we could find a set of cases in which we could say from the outset that there must be some appropriate target, without knowing who that target was would be dysfunctional. We can't know who, if anyone, could be a suitable target of moral education, or whether anyone could be benefitted by blame-based solidarity simply by a description of the outcomes on these scenarios. Instead, we need to establish (as at least Pettit and Copp seem to acknowledge) that there is something other than an individual agent who can be an appropriate target of blame here.

8 Does the Rights-Based Argument Involve a Quantum of Blame Error?

Suppose this diagnosis of a problem in Responsibility Gap arguments for backward-looking responsibility is correct. How does the Wringe-Korsgaard argument for forward-looking collective obligations fare in the light of this discussion?

The argument starts by identifying an obligation which must (allegedly) fall on someone and then searching around for potential obligation bearers, one of whom is preferred on somewhat abstract theoretical grounds (namely, non-arbitrariness). To that extent, it seems to run parallel to the 3-step argument pattern that we have been discussing. In a step analogous to STEP 1, we identify an obligation with a so-far unspecified bearer. Then, in STEP 2, we eliminate from consideration a range of agents who are candidate obligation-bearers. Finally, in STEP 3, we identify some further

entity – the collective consisting of everyone – and assign it the role of obligation-bearer.

There are some significant disanalogies between the rights-based argument and the three arguments which I criticised in the first part of the chapter. Here is one that is worth mentioning. Pettit, Copp and Chant, described situations in which the only plausible candidates for being bearers of responsibility other than the one they finally settle on are individual agents. In the rights-based arguments we need to show that the obligations correlative to a welfare right don't fall on any particular individual. But we also need to show that they don't fall on certain kinds of collective agent: nations, international organisations and so on. I mention this point to set it aside: whatever its merits, it certainly does not establish that the rights-based argument is more robust than the arguments of Pettit, Copp and Chant.

Here is another disanalogy. In HERALD, BOARD and STAND-OFF we were concerned either directly or derivatively with questions about the appropriate allocation of blame. But insofar as the rights-based argument is based on forward-looking considerations – insofar as its concern with outcomes is a concern with *preventing* states of affairs questions about blame need not arise at all. If all goes well, we won't need to ask questions about who needs to be blamed for bad outcomes, because no-one will need to be blamed at all.

Again, though, we might find this unsatisfactory. For we might wonder why the fact that one set of arguments involves considerations about blame while another does not should make one more robust than the other. However, there is a straightforward answer. The fact that HERALD, BOARD and STAND-OFF involved situations in which questions of blame arose meant that the conclusions which we wished to draw about those situations needed to be answerable to standards governing our blaming practices. This fact did not, of course, establish that there was no responsibility to be assigned in the cases under consideration: someone who thinks, as I do, that these are in fact cases where there is non-distributive collective responsibility will also think, as I do, that they are cases where our practices admit of blame. But, as I pointed out in Section 6, there is nothing about those practices to support the idea that we can start off by assuming that there is blame to be distributed here.

If this is correct, then the arguments of Pettit, Copp and Chant seem vulnerable in a way that need have no parallel in the rights-based argument. For questions about the allocation of blame do not seem to play a significant role in the rights based argument. The key premise in the argument concerns the relationship between rights and obligation: blame is not mentioned, and plays no explicit role in the argument. If so, it is hard to see how the Quantum of Blame objection could get a grip here.

9 Accountability Versus Blame

One might still be suspicious. Even if the rights based argument does not rely explicitly on any considerations about blame – insofar as neither the word ‘blame’ nor any of its cognates appears in any of its premises – it might still rely on that notion implicitly. This would be true, for example, if the notion of obligation which featured in the correlativity thesis needed to be analysed in terms of blame: for example, if one took ‘X is under an obligation to Phi’ to be appropriately analysed as ‘It would be appropriate to blame X if X were to fail to Phi’. It might also be true if the correlativity thesis entailed certain further claims which employed notions which were themselves to be analysed in terms of blame. For example, the kinds of claims about forward-looking obligation to which the correlativity thesis gives rise might entail further claims about backwards-looking responsibility which themselves give rise to claims about blame. In either case, one might think, there might be room for considerations about the nature of appropriate blame to get a grip on the rights-based argument, just as they did on the arguments of Pettit, Copp and Chant.

I shall address these arguments in a moment. However, it is worth saying something about why it is reasonable to be confident that they will fail. The functions of our practices of assigning forward-looking responsibility mean that they are, and should be, subject to different kinds of pressures from our practices of assigning backward-looking responsibility. Section 6 suggested that the functions of blame included moral education and the expression of solidarity with victims of wrongdoing. When a bad outcome has occurred, there may be various reasons, internal to our practice of blame, but intelligibly related to the functions of those practices, which explain why

no-one should be blamed. If the bad outcome is the result of an action of someone too young to be able to reasonably foresee the consequences of their actions, or to be capable of coming to foresee them in the light of the blame heaped upon them, or of someone incapable of taking those consequences into account, or of a freak accident, we might find blame inappropriate.

However, the functions of ascription of forward-looking responsibility are different from the functions of ascribing blame. One function of such ascriptions is the assignment of the (prospective) costs of preventing various kinds of bad outcome. When a bad outcome is in prospect, we cannot be as sanguine about allocating forward-looking responsibility as we might be when allocating blame for something bad that has already happened. To refuse to allocate costs in such a case is to decide, by default, that they will fall in a particular way. (If the bad outcome is someone's being deprived of the means of subsistence, the default bad outcome which occurs if we fail to allocate the costs of preventing the outcome is that they are in fact deprived.) Of course, it may be that by some stroke of good fortune a bad outcome that we took to be in prospect turns out not to be: someone might dodge a (literal or figurative) bullet or become the recipient of manna from heaven. But such cases are not parallel to those where our practices dictate that blame is not to be assigned for a bad outcome: they are parallel to cases where there turns out to be no bad outcome for which we need to allocate blame.

It is tempting to put this point by saying that although there is no reason for thinking that there is a fixed quantum of blame to be assigned, there is a fixed quantum of preventive cost. But this is not quite correct. There may be more and less efficient ways of allocating the costs of preventing various kinds of bad outcome, and facts of this sort may make a difference to how we allocate costs. The point is rather that once it is given that certain kinds of outcome are to be prevented, *something* has been fixed: not something quantitative, but something qualitative. And to say that people have welfare rights is to say that those outcomes are to be prevented. The costs of achieving that goal aren't up to us in the way that the question of whom to blame when something goes wrong – for example, our failing to achieve it might be. (It is up to us whether or not we choose to take up any given goal, of course: but whether or not this has any bearing on the question of what

rights we have is another matter: nothing in what I have said entails that it must be.)

If this is correct there are positive reasons for thinking that the Wringe-Korsgaard argument is not vulnerable to the quantum of blame objection: it is that there are positive reasons for thinking that it is not. The reasons we have for thinking that gaps in backward-looking responsibility are tolerable are not paralleled by reasons for thinking that gaps in the allocation of obligations relating to the realisation of rights will be equally tolerable. When we are considering fundamental rights such as the right to a basic level of subsistence, there is good reason to think such gaps are intolerable. They will leave open the possibility of states of affairs which the right in question was supposed to make impossible.

10 Relating Forward-Looking and Backward-Looking Responsibility

One significant loose end remains: the question of the relationship between forward-looking obligation and backward-looking responsibility. Here is a natural view of this relationship.

> (NV) If one has a forward-looking obligation to phi and one fails to do so, one is responsible for one's failure to phi.

NV seems incompatible with the views about blame and forward-looking responsibility that I have defended here. For those views leave open the possibility of situations where some entity is appropriately assigned a role as the bearer of an obligation correlative to some right, but where that entity is not appropriately blamed when the right is not realised. (A concrete example to help focus intuitions: suppose the Global Collective should be seen as the obligation bearer of last resort for putative rights to subsistence and health care (Wringe 2005; Wringe 2014; Wringe 2018). It might still be pointless – and therefore inappropriate – to blame that collective for failing to fulfil this obligation.)

However, it is not clear we should accept NV. Suppose one thinks that backward-looking obligation and moral blame are related in the way

suggested in Section 5. If so, one might also hold a similar view of the relationship between forward-looking obligation and moral exhortation:

> (H) X has a forward-looking obligation to Phi iff it would be appropriate to exhort X to Phi.

But now consider the following:

> (HB) It is appropriate to blame X for failing to Phi if they failed to Phi and it was appropriate to exhort them to Phi.

HB follows from (H) and (NV). But HB seems implausible: it seems to misrepresent the way in which exhortation and blame are symmetrical. For it seems easy to imagine cases where it would be appropriate to exhort me to behave in a particular way, but inappropriate to blame me if I fail to do so.

Suppose I have promised my co-author that every day for a week I will start my working day by working for an hour on my draft of a co-written paper with a view to having the paper finished on Friday. Suppose that you see me on the way to the office on, say, Wednesday. You might exhort me to start work on the paper as soon as I get to the office. Suppose I get distracted and fail to do so. You learn about this on Thursday evening.

Blame might be inappropriate for a variety of reasons. Maybe I manage to make up the time I missed later on Wednesday. Maybe I manage to complete my version of the draft on Thursday. Maybe I am a sensitive soul suffering so badly from writer's block that even the smallest amount of criticism will lead me into a spiralling crisis of confidence. And so on. We might think that these points are objections to H. But they might also give us reason for doubting NV.

NV's initial plausibility may arise from considerations about reparative duties. Often, when one fails to do what one has an obligation to do, something will need to be done to remedy the situation that has come into existence because of this failure. In many such cases there will be reparative obligations: obligations to do what is needed to repair the situation as far as possible. One might think that reparative duties of this sort should fall, in

the first instance, on those whose failure to do what they are obliged to do have brought the reparative duties into existence.

If you are starving because I have deprived you of access to food, then the onus is, one might think, on me rather than anyone else to restore your access to food. Likewise, plausibly, the costs of so doing. If one thinks of liability to blame as akin to a kind of reparative cost, and also holds that liability to appropriate blame is, or entails backward-looking responsibility, then this line of thought could be used to support something like NV. At any rate, it might do so if we think of there being a fixed amount of liability to blame which must be absorbed by some agent. But to think of liability to blame as a cost in this way is, precisely, to think of it in precisely the way that is implicated in the Quantum of Blame error. Once we have freed ourselves of it, NV will no longer seem to derive support from this source.

In short, NV may be natural, but it is by no means obvious that we need to accept it. Furthermore, its seeming to be true may depend on precisely the error that those who are impressed by the Quantum of Blame Error in arguments for collective responsibility see to draw our attention to. The advocate of forward-looking responsibility need have no qualms about rejecting it.

11 Conclusion

The rights-based argument provides us with a plausible case for thinking that there are non-distributive forward-looking obligations. Unlike the arguments of Pettit, Copp and Chant it does not seem to rely on anything like the ‘Quantum of Blame’ error. Those who think that the idea of non-distributive obligations can or should play a significant role in our moral thinking have good reason to focus on cases where we are concerned with what to say about what people should do, rather than situations where we are trying to assess what they have done or failed to do.

* Previous versions of this paper have been delivered at workshops at Jyväskyla, Bilkent and JNU and in the parallel sessions at the Joint Session of the Aristotelian Society at Birmingham in 2024. I’m grateful for comments I have received on each occasion, some of which I’m aware I have not responded to adequately here, to Rooh Haghshenas for

commentary on the version delivered at Bilkent, to the organisers of each event and in particular to Bhaskarjit Neog.

Notes

1. Although I take this strategy for arguing for backward-looking collective responsibility to be flawed, I am not a sceptic about the existence of backward-looking collective responsibility itself.⏎
2. I'm unsure who originated this evocative label, which I have adopted gratefully.⏎
3. To make a move is a kind of action. We shouldn't be distracted here by the fact that philosophers of action sometimes use the phrase 'bodily movements' as a quasi-technical term to refer to items that are perhaps not actions.⏎
4. In legal contexts, there might be good reasons for a presumption in favour of a particular agent being responsible for a range of outcomes in certain kinds of situation. For example, in cases such as the Herald of Free Enterprise we might think there is a presumption in favour of taking the ship's master or the owner of the ship to be responsible. But even if this is a good idea for administrative purposes, the truth might be otherwise. (We should also distinguish between this possibility, and the idea that in certain kinds of case an individual or group might be presumptively liable to paying compensation for a bad outcome: an obligation to pay compensation can be understood as a forward-looking obligation.)⏎

References

Arruda, Caroline. 2017. ‘How I Learned to Worry About the Spaghetti Western: Collective Responsibility and Collective Agency’, *Analysis*, 77: 249–259.⏎

Bennett, Christopher. 2013. ‘The Expressive Function of Blame’, in D. Justin Coates & Neal A. Tognazzini (eds), *Blame: Its Nature and Norms*, pp. 68–83. New York: Oxford University Press.⏎

Brink, David and Nelkin, Dana. 2013. ‘Fairness and the Architecture of Responsibility’, *Oxford Studies in Agency and Responsibility*, 1: 284–313.⏎

Chant, Sara. 2015. ‘Collective Responsibility in a Hollywood Standoff’, *Thought*, 4(2): 83–92⏎

Coates, Dustin and Tognazzini, Neal. 2012. ‘The Nature and Ethics of Blame’, *Philosophy Compass*, 7(3): 197–207.⏎

Collins, Stephanie. 2016. ‘The Claims and Duties of Socioeconomic Human Rights’, *Philosophical Quarterly*, 66(265): 701–722.⏎

Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.⏎

Collins, Stephanie. 2023. ‘Abilities and Obligations: Lessons from Non-agentive Groups’, *Erkenntnis*, 88(8): 3375–3396.⏎

Copp, David. 2006. ‘On the Agency of Certain Collective Entities: An Argument From “Normative Autonomy”’, *Midwest Studies in Philosophy*, 30(1): 194–221.⏎

Copp, David. 2007. ‘The Collective Moral Autonomy Thesis’, *Journal of Social Philosophy*, 38: 369–388.⏎

Copp, David. 2012. ‘The Collective Moral Autonomy Thesis: Reply to Ludwig and Miller’, *Journal of Social Philosophy*, 43(1): 78–95.⏎

Duff, R. Anthony. 1986. *Trials and Punishments*. Cambridge: Cambridge University Press.⏎

Fricker, Miranda. 2016. ‘What is the Point of Blame? A Paradigm-Based Explanation’, *Noûs*, 50: 165–183⏎

Hindriks, Frank. 2009. ‘Corporate Responsibility and Judgment Aggregation’, *Economics and Philosophy*, 25(2): 161–177.⏎

Korsgaard, Christine. 2018. ‘The Claims of Animals and the Needs of Strangers: Two Cases of Imperfect Right’, *Journal of Practical Ethics*, 6(1): 19–51.⏎

Lawford-Smith, Holly. 2015. ‘What “We”?’ *Journal of Social Ontology*, 1(2): 225–250.⏎

Ludwig, Kirk. 2007. ‘The Argument from Normative Autonomy for Collective Agents’, *Journal of Social Philosophy*, 38(3): 410–427.⏎

McKenna, Michael. 2013. ‘Directed Blame and Conversation’, in D. Justin Coates & Neal A. Tognazzini (eds), *Blame: Its Nature and Norms*, pp. 119–140. New York: Oxford University Press.⏎

MacNamara, Coleen. 2015. ‘Reactive Attitudes as Communicative Entities’, *Philosophy and Phenomenological Research*, 90: 546–569.⏎

McGeer, Victoria. 2013. ‘Civilizing Blame’, in D. Justin Coates and Neal A. Tognazzini (eds), *Blame: Its Nature and Norms*, pp. 162–188. New York: Oxford University Press.⏎

Miller, Seumas. 2007. ‘Against the Collective Moral Autonomy Thesis’, *Journal of Social Philosophy*, 38(3): 389–409.⏎

O’Neill Onora. 1986. ‘Hunger, Needs and Rights’, in S. Luper-Foy (ed.), *Problems of International Ethics*, London: Westview. Reprinted as 1988. ‘Rights Obligations and Needs’, in G. Brock (ed.), *Necessary Goods: Our Responsibilities to Meet Others’ Needs*, Lanham: Rowman and Littlefield.⏎

Pettit, Philip. 2007. ‘Responsibility Incorporated’, *Ethics*, 117(2): 171–201.⏎

Pinkert, Felix. 2014. ‘What We Together Can (Be Required to) Do’, *Midwest Studies in Philosophy*, 38(1): 187–202.⏎

Pogge, Thomas. 1992. ‘O’Neill on Rights and Duties’, *Grazer Philosophische Studien*, 43(1): 233–47.⏎

Schwenkenbecher, Anne. 2021a. *Getting Our Act Together: A Theory of Collective Moral Obligations*. New York: Routledge.⏎

Schwenkenbecher, Anne. 2021b. ‘Structural Injustice and Massively Shared Obligations’, *Journal of Applied Philosophy*, 38(1): 1–16.⏎

Schwenkenbecher, Anne. 2022. ‘Collective Inaction, Omission, and Non-action: When Not Acting is Indeed on “Us”’, *Synthese*, 200(5): 1–19.⏎

Shoemaker, David. 2015. *Responsibility From the Margins*. Oxford: Oxford University Press.⏎

Shoemaker, David. 2017. ‘Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame’,

Philosophical Review, 126(4): 481–527.⏎

Srinivasan, Gopal. 2012. ‘A Human Right to Health? Some Inconclusive Scepticism’, *Aristotelian Society Supplementary Volume*, 86(1): 239–265.⏎

Sen, Amartya. 2004. ‘Towards Elements of a Theory of Human Rights’, *Philosophy and Public Affairs*, 32(4): 315–356.⏎

Strawson, Peter. 1962. ‘Freedom and Resentment’, *Proceedings of the British Academy*, 48: 187–211.⏎

Tognazzini, Neal and Dustin, Coates. 2021. ‘Blame’, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition). https://plato.stanford.edu/archives/sum2021/entries/blame/.⏎

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.⏎

Wringe, Bill. 2005. ‘Needs, Rights, and Collective Obligations’, *Royal Institute of Philosophy Supplement*, 57: 187–208.⏎

Wringe, Bill. 2014. ‘From Global Collective Obligations to Institutional Obligations’, *Midwest Studies in Philosophy*, 38(1): 171–186.⏎

Wringe, Bill. 2018. ‘Global Obligations and the Human Right to Health’, in Kendy Hess, Violetta Igneski and Tracy Lynn Isaacs (eds), *Collectivity: Ontology, Ethics, and Social Justice*, pp. 153–175. Lanham MD: Rowman & Littlefield International.⏎

Wringe, Bill. 2020. ‘Global Obligations, Collective Capacities, and “Ought Implies Can”’, *Philosophical Studies*, 177(6): 1523–1538.⏎

OceanofPDF.com

6 Moral Responsibility of Random Collectives

Bhaskarjit Neog

DOI: 10.4324/9781003612827-7

Most moral philosophers tend to think that random collectives are like moral phantoms. For them, such collectives appear to act in a morally relevant way but lack the credentials necessary for being responsible *as* collectives. Their apparent moral agency is merely an illusion, arising from the actions or omissions of individuals who happen to share certain features relevant in a particular context. Thus, when they fail or succeed in acting morally, according to this view, their alleged actions or omissions must be understood in distributivist terms, i.e., by reference to the positions of their constituent individuals.

This view about random collectives enjoys widespread acceptability. Virginia Held, who coined the term random collectives, maintains that when a collection of individuals, let's say, R is 'morally responsible for the failure to do A, then every member of R is morally responsible for the failure to do A, although, perhaps in significantly different proportions' (1970: 471).[1] Larry May, another prominent name in this area, contends that responsibility attribution to random collectives is shared among individuals, even when it cannot be apportioned equally among them (1991: 274). Hans Bernhard Schmid identifies this as a pattern of theoretical commitment where one is bound to believe that 'whatever responsibility random collections might have cannot be genuinely *collective*, but has to be some *distribution of individual responsibility*' (2018: 92).[2] Despite their differences, the underlying view behind these positions is that whatever responsibility is attributed to random collectives, it must ultimately be recoverable in terms of the responsibility of its members.

But how conclusive is this view? Instead of asking whether random collectives are capable of possessing unified agency, what if we ask whether a collection of relevant inactions in a particular situation itself constitutes a morally important phenomenon? In this chapter, I explore the implications of this question. I argue that certain small-scale random collective cases cannot be adequately understood by assuming that their relevant moral obligation is either individually distributed or jointly performed. In particular, I focus on cases in which a moral demand confronts a group of similarly situated individuals but can be discharged by any one of them. I argue that such cases reveal a distinctive possibility overlooked by standard distributive accounts, namely, that an obligation may be collectively held while being individually dischargeable. I call the action through which one member discharges such a collectively held obligation a quasi-representative action.

The key thought behind my proposed position is that certain random collective contexts generate morally relevant outcomes for which individual members cannot be held morally responsible, either individually or jointly. Constituent members cannot, in the relevant sense, be held responsible individually merely by virtue of their separate failures, because those failures do not, taken separately, exhaust the normative structure of the outcome. Nor can they be held responsible jointly, since the action concerned does not require a typical joint effort. Nonetheless, when someone spontaneously discharges the responsibility by treating it as a matter of collective responsibility, the resulting moral benefit accrues to all members of the collective, irrespective of whether we take responsibility to be a matter of attributability or accountability.[3] This benefit may be broadly understood as the avoidance of collective-level blame for morally significant inaction.[4] What needs to be explained here is not merely why the rescuer deserves praise, but why blame for the non-discharge of the collective obligation, although separate grounds for personal criticism, may remain. An individualist explanation can explain the first fact, but not the second one, without appealing to a collective normative transformation.

Before going ahead with this agenda, it is important to note that the idea of moral responsibility taken here is intertwined with both its backward- and forward-looking senses. Forward-looking responsibility in its most

basic sense may be understood as taking responsibility for what will happen in the future by acting in the present with a focus on future consequences. If X has a forward-looking responsibility to do P and X fails to discharge it, the failure gives rise to a corresponding sense of backward-looking responsibility. This, however, need not necessarily mean that X is thereby necessarily blameworthy for the failure. Second, I do not claim that random collectives have an agency of their own in any reasonable sense. What I suggest is that, in certain small-scale cases, their moral standing can be transformed by the way a shared obligation is discharged. In such cases, responsibility attribution tracks a change in the normative stance of the whole collective. The claim is not that the collective possesses agency in the manner a corporate or a committee does. Rather, the claim is that normative subjecthood and agency may come apart in these cases. Thus, the issue in question is whether an individualist explanation suffices to account for this transformation, or it requires an explanatory approach that recognises a non-distributive form of collective responsibility.

The structure of the chapter is as follows. Section 1 examines two influential individualist interpretations of moral responsibility in random collective contexts and argues that, despite their initial appeals, both are inadequate for explaining the kind of issues that are at stake in certain small-scale cases. Section 2 introduces an unorthodox random collective case with three different versions, each of which successively brings to light a form of moral failure and success that resists standard distributivist analysis. Section 3 critically analyses the most promising individualist response to the unorthodox case and argues that it is nonetheless a radically insufficient explanation. Section 4 develops a non-individualist account based on the concept of quasi-representative action. It also shows how such actions can be attributed to a group without presupposing collective agency. Section 5 elaborates this account by highlighting its constitutive features and responds to potential objections. Section 6 concludes by situating the position within a historical context and drawing out its broader implications.

1 Moral Responsibility and Random Collectives

Our everyday life offers countless situations in which strangers should join hands to accomplish things from a moral perspective (Hindriks 2019).[5] However, not all moral imperatives of joining hands demand large-scale involvement. Many operate within smaller, localised contexts where particular groups must act responsibly. For instance, the onlookers of a racially motivated abuse in a public gathering, the bystanders who stand as mute spectators of a hit-and-run accident, and the beachgoers who behave indifferently when a child drowns in front of their eyes. When such groups or collectives fail to discharge these obligations from a moral point of view, we naturally make statements about their failures. We hold them morally accountable for their lack of moral concerns and compassion.

This raises a deep conceptual issue. Moral philosophers have long maintained that genuine moral responsibility can be attached only to entities that possess agential capacities. Groups or collectives of a certain kind, such as corporate organisations, committees, universities, and sports teams, are considered capable of exhibiting agential capacities like individual human persons. They bear these capacities by virtue of their having a particular structure or organisational mechanism. Random collections of individuals, by contrast, lack these structural features and capacities. Without an organised mechanism for deliberation, random collectives cannot formulate collective decisions or reach a unanimous conclusion about what is good and what is bad.[6] Nor can they appropriately channel their constituent members' thoughts and intentions in response to a given situation.[7] Given this, it looks both odd and improper to make moral statements about random collectives. How should we then make sense of ordinary moral assessment of such groups?

There are two familiar ways of explaining the moral assessment of random collectives: First, interpreting the alleged responsibility in terms of the individual responsibility of the constituent members on enumerative grounds; second, interpreting it as a matter of their joint responsibility on the ground of joint realizability.[8] These two interpretations may look attractive on a number of grounds. First, constituent individuals, unlike their collectives, have agency of their own, with respect to the action or omission in question in terms of their potential contribution. Second, individuals will continue to be morally responsible for their individual actions and

outcomes, which fall directly under their individual control, regardless of the normative force of the context in which they are part. Third, nothing can be done at the level of the collective if no individual is ready to move their head to make a difference.

According to the first interpretation, when *all* the individual members have the responsibility to do X, *each one* bears individual responsibility for doing X at their personal level. This view is grounded in the thought that the collective can discharge responsibilities only through the aggregation of individual actions. The difficulty with this interpretation, however, is that it fails to adhere to a very crucial principle of morality, namely, the *ought implies can*.[9] According to this principle, an agent can be under an obligation to perform an action only if that action is within the agent's power to perform. In cases where X requires the coordinated involvement of multiple individuals, no single individual is individually capable of performing X on their own. Consider the above small-scale examples of the bystanders, onlookers, and beachgoers. In each case, the situation typically requires the coordinated efforts of multiple individuals. No individual can, acting alone, do what is required to be done in such a situation. Thus, no individual can be held responsible for the death of the victim as such. They may, at most, be individually responsible for what they could have done at their personal level. But being responsible for one's failure to perform contributory actions is different from being responsible for the overall outcome.

The second interpretation escapes this difficulty as it finds its rationale in the jointness of the desired action. On this view, when individuals are capable of *jointly* doing X, they bear *joint responsibility* to do X.[10] Each individual member's responsibility is thus intertwined with that of other members, since X requires the joint involvement of multiple individuals. Unlike the first interpretation, this account seems to satisfy the *ought implies can* principle: the doability of X does not reside in any single person's capacity, but in the joint force of all the individual members acting together. Individuals are aware of their joint capacity, and they believe that others also have similar awareness that they ought to do X.[11] The possibility of this joint responsibility, however, does not obliterate their moral responsibility on the personal front. Individuals will continue to bear moral

responsibility at their personal level for their contributory actions that are under their control.

An important feature of these two interpretations is that they both require cases where the desired moral outcome can be achieved only through the participation of many, or most, individuals. In each case, the moral evaluation proceeds by reference to collective successes—outcomes that require coordinated effort and widespread active participation. But this invites a critical question. Why does the desired moral outcome require widespread active participation? Are there no situations in which a morally desirable outcome can be brought about by the effort of just one individual, even though many are present?

I argue that such cases do indeed exist, especially in the context of smaller collectives. In such cases, the moral demand of the situation can be met by any constitutive member. This, however, must not be confused with the individuation of the obligation itself. The fact that any individual member can fulfil the demand does not mean that the demand is separately assigned to each member. When one individual discharges this obligation, the resulting moral benefit accrues to the rest of the group. This benefit may be understood in terms of no longer being blameworthy for the inaction or failure.

This brings us to an important clarificatory issue regarding the object of moral responsibility in a collective context. An object of responsibility is a moral imperative that an agent ought to do or discharge in a given situation. The question is whether, in a collective context, such an object must always involve an action that is jointly or collectively produced. Most writers on collective moral responsibility answer this question in the affirmative. I am inclined to resist this affirmation. A collection of individuals may face a moral imperative without that imperative being necessarily an action that is achievable only through a typically coordinated collective action. Each individual of the collective may equally experience the normative force of the imperative, even though coordination with others is not required for its fulfilment. This is because, prior to becoming the responsibility of any particular individual, it is an object of responsibility for all.

This distinction also demands that we separate the bearer of an obligation from the agent of discharging it. The fact that each individual is capable of performing the required action does not necessarily mean that

each of them also bears the obligation in an exclusive sense. There are situations where the obligation may be collectively held while remaining individually dischargeable- meaning that any suitably positioned member can satisfy the obligation directed at the collective. So, what matters most is not merely who can perform the action but to whom the moral demand is normatively directed.

It is important to note, before proceeding further, that even though random collectives lack formal organizational structure, certain situations can generate morally relevant relations among their members. Their moral significance comes to light when we attend to the ways in which individuals are interconnected and interdependent in facing a particular moral imperative. Moreover, random collections of people should not be assessed by indiscriminately applying universal moral standards to their constituent individuals. This not only makes the moral imperative coming out of the crisis situation obscure, but also gives a one-sided understanding of human behaviour in public spaces. It is thus important to attend closely to how such collectives comport themselves in specific crises or emergency situations.

2 An Unorthodox Random Collective Case

Consider the following small-scale case:

> An elderly person gets trapped in a manhole during a sudden flash flood in a city suburb. A group of bystanders witnessed this unfortunate event unfold. The physically weak person is on the verge of losing his life as he will soon be swept away by the water if not pulled out quickly by someone. Consider further that it takes just one person to give the necessary push to the elderly person so that he can lift himself up with his upper body. But nobody comes forward. As a result, the elderly person is swept away and dies.[12]

There could be several reasons for this death, without any single individual being the cause. Nevertheless, the fact that the person dies is morally significant in relation to a putative collective—the collective that did not act but could have acted had it been responsive to the situation.[13] Each individual present there could have rescued the elderly person, yet they all failed to intervene. It is tempting to treat the group as an appropriate object of moral assessment. But what grounds such an assessment?

To shed further light on this, let us consider a second version of the case. With everything else remaining the same, except that this time a soft-hearted bystander comes forward to rescue the victim. As she is about to pull him out, she realises that other individuals remain completely indifferent to the incident. They could easily come forward to assist her, but they do not.[14] This indifference upsets her, and she subsequently backtracks. As a result, the elderly man is swept away and dies. If we are to ask who is responsible for this death, the most obvious answer would be: 'all of them'. Although this formulation of 'all of them' might appear to reduce responsibility to each individual taken separately, there is a substantial sense in which 'all of them' should instead be understood as 'all of them taken together'. In this latter sense, each individual is connected to the others in facing the common moral demand of rescuing the person. Their agential interconnection becomes apparent even though the moral demand does not necessarily require the participation of all or most members of the group.[15] Each individual faces the demand because it arises as a unified moral imperative. Before it becomes anyone's particular duty, it exists as the duty of *all*. The death does not result merely from a series of independent inactions, but from a single inaction or failure attributable to the putative collective as such.

To strengthen the point even further, consider a third version, this time with a positive outcome.[16] Unlike the previous two versions, where the elderly person dies, here a kind-hearted bystander promptly takes it upon herself to rescue him. The other individuals could assist her, but there is no need since her own efforts are sufficient to rescue the victim. The fact that her individual action is sufficient for the rescue does not imply that the underlying moral demand exclusively belongs to her. The demand remains as one that confronts the members collectively, even though its fulfilment

does not require their exclusively unitary collective effort. The other individuals observe the rescue effort, but make no contribution to the outcome. At this stage, this might appear to involve a form of joint action. But a careful observation reveals that the act falls short of having the property of jointness, since it does not require the *joining of hands*. Nevertheless, the kind-hearted bystander performs the act with the others in mind. She performs it as if it were the act of all. This distinguishes the act from being a purely individual act. There is a difference between doing something morally desirable on one's own and doing the same in a context where others are subject to the very same moral demand. The latter differs from the former in that it is not performed from a standpoint characteristic of ordinary individual moral action.[17] Rather, it is performed from a shared moral perspective in which the obligation is common to all.

This unorthodox case throws interesting light on several important dynamics of random collectives. First, it demonstrates that the moral duty performed in a collective context may sometimes be located in a normative space accessible to any member of the collective. Second, inactions or failures in a collective context need not always be assessed against the assumption that morally required action must involve the explicit participation of many or most individuals with rationally acquired coordination. Third, such cases of collective contexts shift our attention away from standard textbook examples where obligation is essentially either a joint obligation or the obligation of a collective agent.

This brings us to a crucial question: can an explanation of individualist orientation adequately account for these unorthodox situations? I am inclined to argue that the answer is negative. I suggest that we look beyond individualist readings and embrace a more nuanced non-individualist reading. But before pursuing this alternative, let us, however, examine why and in what respects individualist readings fall short.

3 Individualist Interpretation and the Standard Analysis

Since random collectives lack their own agency, philosophers naturally look for an individual-level analysis to resolve the question of their alleged

moral responsibility. Both the interpretations discussed in the second section make individual agency the focus of their explanations. When tested against the ‘ought implies can’ principle, however, only one of them survives critical scrutiny. This is because only one of them places the agency issue at the joint level. Since the morally desired action requires the involvement of multiple individuals, joint-level coordination between members can satisfy the normative demand that no single person could meet alone. Yet this interpretation faces a significant difficulty. It becomes redundant, given that the desired action here does not necessarily require multiple agents. A spontaneous response from a single individual would suffice to accomplish the goal.

Does this mean we should return to the first interpretation and examine how well it goes while addressing the unorthodox case, especially its third version? I argue that this interpretation also fails. Let us examine this in more detail. A thoroughgoing individualist interpretation can be cashed out in two possible ways: first, by distributing the responsibility equally among all members of the collective, and second, by assigning responsibility only to those who possess the relevant capacities (the ability to rescue the elderly person).

The first alternative that attributes responsibility to every individual presumes that each has the capability to fulfil the obligation. This satisfies the principle of *ought implies can*. A supporter of this might therefore think that the third version does not create any difficulty. Each member has an obligation to rescue, and each is capable of doing so, the fact that one member does it eventually merely explains why the harmful outcome is avoided. Since each individual is individually capable of discharging the obligation, this view takes each to be individually responsible for the wrongful state. Defenders of this view might respond that non-acting members continue to bear responsibility for their inaction, even though the obligation has already been discharged by someone else. The problem with this response, however, is that it assumes rather than establishes the distributive character of the obligation. The fact that each individual was capable of doing the relevant action shows that each could have done it; it does not yet show that each separately bore the obligation.

Another reason for attributing moral responsibility to individuals is the assumption that an individual member’s individual failure is considered to

be sufficient for causing the wrongful outcome. But this argument is equally problematic, for the same causal relations apply symmetrically to all other members. The crucial issue is not that no individual failure is causally sufficient for harmful outcome, rather, it is that causal sufficiency does not by itself determine the normative bearer of the obligation. Now let's examine the second alternative. On this view, moral responsibility arising from random collective cases should be assigned only to certain members. But who falls under this category of *some*? In our elderly person case, *some* might encompass various candidates who could have contributed to preventing the harm. More specifically, *some* refers to members who share relevant common characteristics that render them suitable bearers of moral responsibility.[18] First, an individual with previous experiences in similar emergencies; second, an able-bodied individual who witnessed the entire incident from beginning to end and yet did nothing without a justifiable reason; and third, an individual who could have carried out the rescue most efficiently without causing any harm or inconvenience to anyone.[19]

If this much is true, this alternative interpretation enters into an uncertain territory. We cannot reliably determine how one should rank the contributory merits of their respective inactions. If one of them is held responsible on one ground, the other two can also be held responsible on two other equally compelling grounds. If previous experience is the criterion that renders the first individual morally duty-bound, why should the second individual, who saw it through and yet did nothing to save that person, be exempt from responsibility? Likewise, if efficiency or the ability to intervene safely is taken to be the decisive criterion, why should the third individual, who could have ensured the near-100-percent safety of the manhole-trapped person, not be considered responsible? The crux of the problem is that no single individual can be said to be exclusively responsible for the death of the person. All of them are responsible.[20] All of them are *collectively* responsible for the failure, as it occurred at the same time. Put differently, since there is no compelling reason why any particular member alone should have saved the elderly person, and since many or all were roughly equally positioned to help, fairness demands that we hold all of them collectively responsible. To blame only some or a few would be unfair.[21]

To sum up the criticism, even though a straightforward explanation of individualist orientation looks promising for the unorthodox case at first blush, it eventually misses a crucial social aspect of interconnectedness. This happens because of the overemphasis on the isolated individual contributions towards the outcome. The account evaluates the situation by completely ignoring the interdependent nature of the situation while unduly prioritising the sufficiency condition of each individual's expected and potential contribution. While individuals may be judged by examining their potential capacities, it is inappropriate to consider these capacities in isolation from their embedded social aspects altogether. The key issue is that moral responsibility in collective contexts cannot sufficiently be understood in atomistic terms. In situations like these, the failure or success of an individual or a collection of individuals needs to be understood by keeping in view the larger context. To properly appreciate these contextual dimensions, it is important that we interpret our desired moral duties and obligations on representative grounds.

4 Possibility of a Non-Individualist Interpretation

The upshot of the preceding section is that a purely distributive interpretation—whether understood in terms of individual responsibility or joint responsibility—does not fully capture the normative structure of certain small-scale random collectives. Such cases call for an alternative, non-individualist reading that is sensitive to the representative feature of positive versions of these cases. How, then, should one develop such an approach? I think the most promising way to develop this is to recognise a category of actions that has a representative character.

Let me clarify what I mean by this representative character. We often act on behalf of others—sometimes voluntarily, sometimes in response to an agent's expectation, and sometimes in fulfilment of an explicit promise. An action counts as representative when it is performed on another's behalf through an agent's acceptance of the interests of that individual or entity. The crucial feature of representative actions is that the agent performing the action does not necessarily have to share the very same interest as those on whose behalf the action is performed. There can be a clear disjunction

between the agent's interests and the interests of the party that they represent.

However, this disjunction is not present in every instance of representative action. Consider again the kind-hearted bystander in the third version of the unorthodox case who rescues the elderly person. Her act of rescue is grounded in her own moral obligation, and yet it is not merely the expression of discharging personal obligation. It also embodies the action that the other individuals are similarly situated to perform in response to the common moral demand. The kind-hearted person operates in collective contexts where others share a similar moral interest in rescuing the victim. Since other individuals face the moral obligation in the same way that the kind-hearted individual does, the latter's intervention functions as a representative moral act on behalf of all. In such a situation, her action transcends the boundaries of a mere individual action and agency, and comes to represent, or gives expression to, a we-perspective shared by most members of the collective. Her action can thus be understood as the individual discharge of an obligation that is held collectively. And it is this combination of collective holding and individual discharge that gives the action a unique status.

It is important to distinguish this we-perspective from the we-perspective found in standard accounts of joint action or shared intention. In the cases considered here, agents do not form a collective intention nor coordinate a behaviour through mutual responsiveness or meshing of sub-plans. The morally desirable actions are performed independently, often spontaneously, without prior awareness of others' positions. Nevertheless, they may be characterised as we-perspectives in the sense that the agent's action is taken as discharging an obligation that belongs to the group as a whole. This perspective is normative and not typically agential because the action is understood in relation to a particular shared situation rather than as a product of collective deliberation.

I propose to call these quasi-representative actions. I call them so in order to ensure their difference from typical representative actions in which an agent acts as a surrogate for another individual or entity. The source of a quasi-representative action arises from a genuine convergence between the individual agent's perspective and the normative perspective shared by the collective. These acts constitute an authentic moral expression shared by all

through individual agency. In this sense, the agent genuinely acts on their own while simultaneously expressing their deeper moral commitments towards the collective.

We can explain this point more clearly in the following way. Suppose that, unlike the situations in those three versions, floodwater is now carrying the elderly person down the street.[22] People at the top end of the street have a chance to help, but they do nothing, and the victim is carried too far away from them. Fortunately, someone at the lower end of the street grabs the elderly person and saves him. Now, is this rescuer's action representative? I argue that it is not. This action lacks a representative or quasi-representative character because the situation fails to generate a we-perspective in the face of the moral demand. For individual members to form a relevant we-perspective in a relevant way, they need to find themselves in close physical proximity (or in face-to-face relation) within the same moral context.

David Miller eloquently brings to light this physical proximity factor while analysing the moral significance of a duty of rescue (2001). He suggests that physical proximity creates relationships between the rescuer and the victim on the one hand, and among the potential rescuers on the other. Such proximity enables them to develop a shared understanding of how they might collectively respond to a common moral demand. In the above example, even though the people at the top end may each share a we-intention, this perspective fails to generate a special normative bond among them. It does not give rise to a kind of normativity relevant to the demand of rescuing the victim. The fact that the person at the other end of the street has ultimately come forward and rescued the victim cannot be attributed to those at the top end. The duty discharged in that case is understood as an individual moral action, not as one that expresses an authentic we-perspective of those people located elsewhere.

By contrast, the third version, the kind-hearted bystander version, does exhibit a genuine we-perspective among the individuals in the face of the moral demand. This perspective arises from the co-members' physical proximity and mutual awareness of one another's responsiveness to the situation. Although these members are not together because of any prior personal goals or objectives, the elderly person's predicament generates in

them an altruistic disposition towards the rescue. Each individual is aware not only of the moral demand but also of others' ability to discharge it. This mutual awareness supports their thought that someone among them ought to discharge it. This includes their realisation that they are not competitors but potential collaborators in discharging this, even though no coordination is required. Given these situational conditions, the action can be appropriately described from a normative we-perspective. I use 'we-perspective' in a weaker normative sense than we-perspective of we-intention standardly used in accounts of joint actions. My sense of we-perspective does not require the formation of a collective intention. It just refers to the way an individual understands the moral demand as one confronting the members of the group as such. Actions performed from such a perspective are not grounded in formal confirmation or agreement. Nor do they rest on mutual promises between them. They are a spontaneous response to an urgent situation and thus are carried out without deliberative coordination among the members.

The authorship of quasi-representative actions bears a distinctive feature. Unlike individual actions rooted in individual agency, quasi-representative actions emerge from a representative and non-individualist foundation that has the tacit recognition of all or most. Thus, for this reason, the agent performing the action does not really own these actions. One might argue that since such actions are ultimately performed by individuals, it is the performing individual who should be recognised as the author. But this inference is misleading. Quasi-representative actions are not performed from an I-perspective—the perspective that marks the identity of a paradigmatic individual action. Rather, they are performed from a shared moral standpoint in which it is important that someone ought to go and act now. This demand is intelligible only when we look at the scenario from a non-individualist moral outlook.

This is a form of awareness that closely resembles what Hans Bernhard Schmid calls a plural pre-reflective self-awareness (Schmid 2018). The heart of Schmid's concept is the idea of 'sense of us' among the individuals. It is this 'sense of us' that drives the individual participants to identify as 'us'. In the present context, the agent acts under a 'we-perspective', rather than an 'I-attitude' typical of individual agency. This explains why the moral effects of such actions go beyond the acting individual. Once a quasi-

representative action is performed, its positive moral consequences go to the rest of the individual members. These members are no longer appropriate targets of collective blame for the morally crucial failure. Instead, they now benefit from having the morally required action completed. Subsequently, the execution of the obligation effectively redeems the moral stake of the entire collection of individuals, which would have been subject to collective moral criticism.

At this point, one might ask about an important question regarding membership in such a collective. Who qualifies as a member whose moral standing is affected? In the case of a random collective, the grounds for membership are rather fluid. Unlike structured collectives, random collectives lack formal criteria—virtually anyone present at the scene may count as a member. This, however, does not imply that every member's moral standing is automatically redeemed by a quasi-representative action. The fundamental condition of membership, I suggest, is proximity. Physical proximity generates a distinct form of moral spontaneity and immediacy. When one finds oneself near someone in urgent need of assistance, one is disposed to respond without first considering whether others present in the same situation will do the same. Sharing the same circumstances also means sharing the same basic capacity to intervene effectively.

That said, proximity alone is not enough. There may be cases in which an individual, though physically present, lacks the capacity to intervene because of physical limitations. In such a situation, the individual is exempt from the primary obligation. Nevertheless, this does not mean that there is no obligation as such. They may still have an obligation, a derivative obligation, such as informing the police or alerting someone better positioned to assist.[23]

One might still question whether the individual who executes the quasi-representative action genuinely acts from a we-perspective. The individual might insist that her intervention is motivated solely by her own altruistic impulses and not by any collective outlook. I am inclined, however, to think that a reflective moral agent would find it implausible to claim full moral authorship of such an action by dismissing the situational relevance of others who face the same moral demand. One may, of course, legitimately claim authorship of a courageous act—especially where others lack the

courage to act. But there is a clear conceptual difference between a courageous act and a representative act. A courageous action typically involves the exercise of exceptional abilities that others do not have. A quasi-representative action, by contrast, does not presuppose superior ability. It arises from occupying a shared moral situation and promptly responding to it on behalf of all who are equally positioned to do so.

5 An Account of Quasi-representative Actions

At first blush, the notion of quasi-representative actions may appear to resemble Kirk Ludwig's account of proxy agency. Ludwig defines proxy agency as a case in which 'one person or subgroup's doing something counts as or constitutes or is recognised as (tantamount to) another person or group's doing something' (Ludwig 2013: 76). His proxy agency involves action types that are inherently representative in nature, as these acts are performed on others' behalf. Take for instance a hired spokesperson for a political party who clarifies the party's stance against corrupt political funding. This clarification constitutes an act of proxy—a surrogate act performed on behalf of the political party. The spokesperson personally may or may not agree with the party's position on funding, which is a different issue. As long as it does not interfere with the execution of the proxy act, the spokesperson's personal beliefs are irrelevant.

Quasi-representative actions, on the other hand, operate differently. Here, individual-level involvement cannot be dismissed or ignored altogether. In fact, individual-level involvement provides the anchoring force that drives others' interest forward. An agent performs a quasi-representative act by way of acting on one's own interest that aligns strongly with the interest of others in which they are a part. While this involves an act of representing things for others, it also represents a deeper normative fact—i.e., the authentic convergence of individual and collective commitment in response to a common moral demand. It is this convergence that makes quasi-representative actions different from proxy agency.

An important reason why we need to recognise quasi-representative acts as a distinct category of action is their spontaneous nature. Spontaneity sets these acts apart from other actions performed in a collective context, such

as joint actions, plural actions, or group actions. Standard analyses of these actions are typically characterised by requiring a binding element—some form of coordination or mutual agreement among the co-members. By spontaneity, I mean the ability to act impromptu, without any premeditated coordination or expressed mutual willingness. This ability to act instantaneously is not just limited to the performer or sub-group alone; it is something that can be attributed as a state of affairs to the whole collective. In circumstances of urgent need, the co-members' individual interests spontaneously converge on recognising a we-perspective, even in the absence of deliberation, communication or agreement.

With this brief overview of representative actions in place, let me now turn to highlighting a few important aspects in concrete terms:

First, quasi-representative actions function as a moral 'vector' that represents the desires and interests of all co-members within a specific context.[24] They speak *for* the entire collection of individuals rather than only for those who actually perform them. While in our ordinary moral practice, the performer in a collective context typically receives the moral praise individually for doing the right thing, quasi-representative actions exhibit an inherent collective character that warrants socialising the positive collective responsibility generated in that context.[25]

Second, quasi-representative actions are performed with a we-perspective that represents the putative group's interest in the situation concerned. This we-perspective is formed because (1) the co-members find themselves in physical proximity with each other, (2) they face a particular common moral demand, and (3) each individual recognises the doability of the act, which causes spontaneity in them to have a particular attribution towards them. All these conditions together generate a form of collective sense of spontaneity among the members.

Third, unlike paradigmatic individual actions, where the agent is the sole owner and author, quasi-representative actions do not belong exclusively to the performing agents. There is an inherent openness in their constitution—an openness that refers to the possibility that others could join in or take over the action. This suggests that no single individual bears an overarching obligation in an exclusively individual capacity; rather, all members share this obligation collectively, even though any suitably positioned member

may discharge it. In this sense, all the members stand on equal moral footing with respect to the imperative or action.

Fourth, individual members share a mutual awareness of their general moral obligation towards the situation. This obligation is primarily on the putative group as a collection. Such an idea of a group emerges when all the individuals find themselves on common moral ground with close proximity to each other. While each individual is capable of fulfilling this obligation, they collectively recognise that it belongs to all of them, equally but non-distributively. This shared awareness of the obligation is constitutive of many considerations: for instance, recognition of foreseeable consequences in case of inaction, concerns about potential conflicts if they all attempt to fulfil it, appreciation of situational urgency, and the like.

Finally, when such actions are performed, the remaining individuals may experience a sense of relief. Where such relief is experienced, it may stem from the fact that the morally desirable obligation has now been discharged by *one of* them. This sense of relief may also stem from their tacit recognition of the fact that they are no longer the subject of moral blame or criticism. The nature of this relief phenomenologically differs from what others who are not part of the collective might feel after seeing or hearing about the whole episode.

Before wrapping up, let me address a couple of potential criticisms. Critics might argue that quasi-representative actions are just a form of individual moral actions occurring within collective settings. Since most moral performances involve some degree of shared we-perspectives, the criticism continues, quasi-representative actions collapse into cases of courageous individual actions in a collective context.[26] This objection, however, does not stand on solid ground. The actions under consideration here are not actions of exceptional individual capacities. Even though such actions are performed in a unique situation, to characterise them as acts of individual bravery or courage would be misleading. They are as much acts of moral appreciation as they are acts of representation. Since they are performed against the backdrop of a we-perspective held by co-members, they have an inherent social constitution that warrants individual-transcending moral attribution.

A second criticism may be raised with respect to the exact moral status of the non-performing members in the collective. One might suggest that non-performing members ought to experience guilt for failing to take it upon themselves and allowing others to perform on their behalf. This guilt, one might add, reveals a form of factual moral failure for not performing the duty. To address this, I suggest that quasi-representative actions undermine the appropriateness of collective guilt for the failure, even though the individuals may continue to experience regret or criticism at the personal level. What may be experienced as guilt can best be understood as a form of personal disappointment—a regret at not having been the one to respond. In the strict sense relevant here, these attitudes need not be understood as expressions of continuing moral failure. They reflect a missed opportunity for moral initiative and not a failure of moral duty.

6 Concluding Remarks

The term random collective was first introduced by Virginia Held in a 1970 chapter, long before the contemporary fields of collective responsibility and social ontology had emerged. Held discussed the issue of the moral responsibility of non-agent collectives using a number of examples. Held's cases analyse situations where individuals who happen to find themselves together, without any prior plan or purpose, face statements of collective responsibility. These cases show remarkable similarity with the infamous case of the murder of Kitty Genovese. Although it is difficult to establish whether Held drew inspiration from the widespread discussion surrounding this tragedy, the timing of her publication just after the incident suggests that this may well be the case.

Kitty Genovese was a young bar manager. She was stalked, raped, and then murdered by Winston Mosley outside her apartment building on 31 March 1964. The New York Times initially reported that nearly 38 witnesses observed the attack but did nothing to stop it. The case generated significant media attention because it exposed critical concerns about our moral and psychological frameworks. Central to these concerns was the issue of the moral responsibility of those people as a collective whose failure to intervene led to the death of an innocent woman. The questions

that drew moral philosophers' attention are—whether the responsibility goes to the individuals individually? Or could there be a sense in which their moral responsibility assumes a collective form?

Most contributors to the literature on collective responsibility, including those who take the line of Held, answer this question in distributive terms. According to them, responsibility in such an unstructured or random collective context can be attached only to individuals, either individually or jointly. There is no way one could here talk about collective responsibility in a substantive sense, meaning that the collectives can never be seen as responsibility-bearing entity.

In this chapter, I challenged this assumption by drawing our attention to the fact that such a view is grounded in an assumption that suggests that morally required actions in such contexts essentially require the active and coordinated participation of multiple individuals. I suggested that this assumption narrows the scope of random collective cases quite substantially. There are small-scale situations that do not necessarily require the coordinated capacities of multiple individuals. They can be addressed by any one of their members. I analysed a positive responsibility case of a small-scale random collective, which is characterised by immediacy and proximity. In cases of this kind, a single agent can discharge a moral obligation in a way that transforms the normative standing of the collective with respect to that specific moral failure. When such an obligation is discharged with a we-perspective, the resulting benefit accrues to all the individuals as a whole rather than merely to the acting individual. I characterised the discharge of this responsibility as a matter of performing a quasi-representative action. The key motivation behind this characterisation is that certain moral outcomes appear to be attributable to all members of a random collective, even though the responsibility at issue cannot be adequately attributed to any member jointly or individually. The moral demand operative in such situations is itself a collective demand— one that cannot be adequately reduced to typical individual obligations as such. If this is correct, then the supporters of the standard view will have to admit the limits of their distributivist agenda and reconsider their claim against the alleged phantomhood of random collectives.

Notes

1. However, Held (1970) maintains that even a random collection of individuals can be meaningfully held responsible when it fails to transform itself into an organised group capable of doing the required act. This claim appears confusing because if the collection of individuals is itself the subject of responsibility for this failure, it is unclear why one must still emphasise the distributive nature of that responsibility. Her account also remains unclear on how we are supposed to individuate the action *A* mentioned in the quote.⏎
2. Schmid (2018) himself does not endorse such an individualist viewpoint. Instead, he argues that random collectives can be held morally responsible for their failure on collectivist grounds. He develops this possibility through his notion called *plural pre-reflective self-knowledge*. He suggests that there can be multiple interpretations of a distributive reading such as (1) each individual is responsible to the extent that they could have made the moral difference by taking part in the action, (2) each one is responsible to the extent that they give it a try at least, irrespective of the nature of the outcome, (3) each one is fully responsible for the harm caused, irrespective of their agential abilities. Also see, Schmid (2018).⏎
3. Alexander Leferman (2023) introduces a spontaneity requirement for analysing the phenomenon of acting together. He raises this as a challenge against the standard theories of joint action. For him, to be spontaneously joined is to be immediately joined. My use of the spontaneity requirement, however, is rather different. I consider spontaneity as one of the many aspects of quasi-representative actions.⏎

4. A common assumption is that moral praise and blame are structurally symmetrical. The conditions under which an agent is a target of praise mirror the conditions under which an agent is the target of blame. This symmetry is generally motivated by the thought that it would be arbitrary to hold agents morally accountable for failures while denying corresponding credits for moral success, or vice versa.⏎
5. Hindriks (2019) discusses the different facets and stages of the moral demand, called the duty to *join forces*. Focusing on the issue of collective harm, he identifies the duty to join forces as irreducibly collective.⏎
6. The condition of a 'decision method' or 'decision-making structure' is widely considered to be the benchmark for identifying a collective agent; see, List and Pettit (2011), Corlett (2001), and Chant (2015).⏎

    Schmid (2018) critically examines the conceptual link between decision procedure (his term) and collective agency. He argues that defining agency in terms of decision procedure leads us to capture agency by its effects rather than its principles. Contrasting with individual agency, he argues that in the case of individuals, we do not suggest that so and so has a decision procedure and thus they are agents.
7. They also fall short of several other conditions highlighted in the literature on collective agency and intentionality. They do not have any joint commitment (Gilbert 1990), and they do not have rational unity (Rovane 2004).⏎
8. There is another important approach that has come to light, particularly in the works of Stephanie Collins (2013). Although originally advocated by Virginia Held, it received a full-fledged treatment in Collin's hand. According to Collins, unstructured groups do not as such have any collective duty. Only structured collectives can have such duties. The relevant

duties in unstrucutred or random collectives must be understood in terms of duties of the individual members. She calls such duties collectivisation duties. In this paper, however, I do not discuss this approach because first, it would require a more exhaustive engagement and second, its argumentative thrust is more or less similar to the broader individualist agenda.⏎

9. The principle simply suggests that an agent is morally obliged to do an act (ought) only when it is possible for them to do (can) it. Often associated with Kant, this principle maintains that our inability to act is sufficient for denying our moral obligation. See, Collingridge (1977), and Lawford-Smith (2012).⏎

    The appeal to 'ought implies can' should not be understood as settling the question of who bears an obligation. The principle does not by itself entail that an obligation that is individually dischargeable must therefore be individually borne. In the following case, any suitable member is capable of satisfying the relevant demand, while the obligation may nevertheless be understood as collectively held.

10. Schwenkenbecher (2014) offers a strong defence of this interpretation. Like many others, however, she discusses only those cases that require the genuine participation of multiple individuals simultaneously. She calls this a non-individualist account.⏎
11. See, Schweikard and Pettit (2006).⏎
12. Cases of this nature are different from the cases discussed by Björnsson (2014; 2020) and Hindriks (2018). Hindriks argues for an irreducible responsibility of random collectives. Björnsson (2014), on the other hand, uses his cases to explain what he calls *essentially shared moral obligations*.⏎
13. By putative collective, I mean the idea of the collective that makes the collective what it is. This thought of the collective

exists not only for those who are a part of the collective, but also for those who are not. This notion of a putative collective is thus different from a mere ideal or organised group that the collective has not formed but could have formed in order to do X together; See Held (1970).⏎

14. An important issue that often demands our attention is the cost of discharging such a duty. In most cases, a rescuer may have to bear some cost in attempting to rescue a victim from death or harm. Generally speaking, the cost must be significantly lower than the harm that the victim would suffer if not rescued. However, here I do not address the issue, as it deserves an independent investigation.⏎
15. One might ask whether the soft-hearted individual is now equally morally responsible as the others. I do not think the answer needs to be positive. She does not acquire any special moral status simply because she appears different from the rest. Once she decides to backtrack and joins the rest of the others in allowing the victim to die, she stands on the same moral footing as the rest of the others.⏎
16. One might here even bring in several versions of such a case. A particularly crucial version could be a scenario in which one or two individuals try to assist the victim but ultimately fail to make any meaningful difference. How does one then assess the moral position of such individuals in comparison with others who do nothing at all? Bernhard Schmid (2018) deals with such cases. I agree with Schmid that while such individuals may not be personally blameworthy given that they tried at their individual level, they are still responsible in a collectivist sense in the sense that the collective in which they are a part has ultimately failed to intervene.⏎
17. David Miller (2001) defends a duty of rescue account from a justice point of view that specifically deals with the issue of

the nature and limits of such a duty. He argues that the duty of rescue implies that the victim has a moral right to be rescued by the passerby, and relatedly, that the rescuer has an enforceable obligation to carry out the rescue work.⏎

18. This form of responsibility is distinguishable from what David Miller (2001) calls remedial responsibilities. Remedial responsibilities are a form of responsibility that can be discussed with respect to both individuals and collectives when it comes to the sufferings and deprivation of certain sections of people living in certain parts of today's world. The collective responsibility of random collectives, however, is different from remedial responsibility. Unlike remedial responsibility, which is corrective and reformatory in nature, the collective responsibility of random collectives is contextual.⏎
19. The list of these three conditions is by no means exhaustive. Depending upon the defining features of a given case, one may identify a list of conditions with different forms and content. Also, here one has the freedom to categorise the members based on their contributory merits on both positive and negative grounds.⏎
20. Collins and de Haan (2021) offers three factors that affect the *scope* of an agent's blameworthiness in collective wrongdoing cases—shared intentionality, interpersonal influence, and common knowledge. While I agree with their proposal, I do not think all these can be easily ascertained in the case of the collective inaction scenarios considered here.⏎

    Besides, I deliberately avoid invoking the condition of shared intention here. I consider shared intentions to be a far more loaded concept than what we need here from the members at their epistemological level. I am motivated by a helpful distinction between thick and thin shared intentions made by Roy and Schwenkenbecher; see, Roy and Schwenkenbecher (2019).

21. I am grateful to Olle Blomberg for helping me articulate this point. I am also indebted to him for extensive comments on an earlier version of this chapter.⏎

    This unfairness complaint need not be understood in terms of unequal outcomes. Instead, it should be understood as an asymmetrical moral burden. In situations where agents are symmetrically situated with respect to a moral demand, there will be no special basis for assigning the burden of blame to only some rather than all. To do so will be to treat equivalent positions as normatively distinct without much logic. This symmetry, I believe, offers us justification for attributing moral responsibility to the collective rather than distributing it among individuals.
22. I am grateful to Luke Roelofs for helping me formulate this additional layer with the unorthodox case.⏎
23. An emphasis on physical proximity might mean that we are excluding those situations where physical presence is not necessary. Take, for instance, online communities that raise funds for victims. I think here we must draw a distinction. The immediacy and spontaneity that we see in our random collective cases are generally absent in online communities.⏎
24. See, Petersson (2008). Petersson discusses the concept of vector while engaging with Tännsjö's (2007) concept of metrics.⏎
25. I recognise that there is a qualitative difference between the moral positions of the performing members and the non-performing members. Addressing this distinction requires a separate analysis for which I do not have enough space here.⏎
26. We may here recall Wilfrid Sellars' view regarding the claim of we-attitudes to intersubjective validity. According to Sellars, in we-intentions an individual intends something *as* a member of a group. Such attitudes, he argues, necessarily involve a shared point of view from which members can critically assess

one another's contributions. For a more detailed articulation of this idea, see Koons (2019).⏎

References

Björnsson, G. 2014. 'Essentially Shared Obligations', *Midwest Studies in Philosophy*, 38: 103–120.⏎

Björnsson, G. 2020. 'Collective Responsibility and Collective Obligations Without Collective Moral Agents', in Saba Bazargan-Forward and Deborah Tollefsen (eds), *The Routledge Handbook of Collective Responsibility*, pp. 127–141. NY: Routledge.⏎

Blomberg, O. and Hindriks, F. 2020. 'Collective Responsibility and Acting Together', in Saba Bazargan-Forward and Deborah Tollefsen (eds), *The Routledge Handbook of Collective Responsibility*, pp. 142–154. NY: Routledge.

Chant, S. R. 2015. 'Collective Responsibility in a Hollywood Standoff', *Thought* 4(2), 83–92.⏎

Collins, S. 2013. 'Collective's Duties and Collectivization Duties,' *Australasian Journal of Philosophy*, 91:231-48.⏎

Collins, S. and de Haan, N. 2021. 'Interconnected Blameworthiness', *The Monist*, 104: 195–209.

Collingridge, D. G. 1977. 'Ought Implies Can and Hume's Rule', *Philosophy*, 52(201): 348–351.⏎

Copp, D. 2007. 'The Collective Moral Autonomy Thesis', *Journal of Social Philosophy*, 38: 369–388.

Corlett, A. J. 2001. 'Collective Moral Responsibility', *Journal of Social Philosophy*, 32: 573–584.⏎

Feinberg, J. 1989. ‘Autonomy’, in C. Jonn (ed.), *The Inner Citadel: Essays on Individual Autonomy*, pp. 27–53. Oxford: Oxford University Press.

Gilbert, M. 1990. ‘Walking Together: A Paradigmatic Social Phenomenon’, *Midwest Studies in Philosophy*, 15: 1–14.⏎

Held, V. 1970. ‘Can a Random Collection of Individuals Be Morally Responsible?’, *Journal of Philosophy*, 67: 471–81.⏎

Hindriks, F. 2018. ‘Corporate Agency: Moral and Amoral’, *Dialectica*, 72: 3–23.⏎

Hindriks, F. 2019. ‘The Duty to Join Forces’, *The Monist*, 102: 204–220.⏎

Koons, J. R. 2019. *The Ethics of Wilfrid Sellars*. NY: Routledge.⏎

Lawford-Smith, H. 2012. ‘The Feasibility of Collectives’ Actions’, *Australasian Journal of Philosophy*, 90: 453–467.⏎

Leferman, A. 2023. ‘Joint Action and Spontaneity’, *European Journal of Philosophy*, 33: 167–182.⏎

List, C. and Pettit, P. 2011. *Group Agency: The Possibility, Design and Status of Corporate Agents*. Oxford: Oxford University Press.⏎

Ludwig, K. 2013. ‘Proxy Agency in Collective Action’, *No ûs*, 48(1): 75–105.

May, L. 1991. Symposia Papers: ‘Collective Inaction and Shared Responsibility’, *Noûs*, 24: 269–277.⏎

Miller, D. 2001. ‘Distributing Responsibilities’, *Journal of Political Philosophy*, 9: 453–471.⏎

Nussbaum, C. M. 2007. *Frontiers of Justice: Disability, Nationality, Species of Membership*. Harvard: Harvard University Press.

Petersson, B. 2008. ‘Collective Omissions and Responsibility’, *Philosophical Papers*, 37: 243–261.⏎

Pettit, P. 2003. ‘Groups with Minds of their Own’, in Frederick F. Schmitt (ed.), *Socializing Metaphysics*, pp. 167–193. MD: Rowan and Littlefield.

Rovane, C. 2004. ‘What is an Agent?’, *Synthese*, 140: 181–98.⏎

Roy, O. and Schwenkenbecher, A. 2019. ‘Shared Intentions, Loose Groups and Pooled Knowledge’, *Synthese*, 198: 4523–4541.⏎

Schmid, H. B. 2018. ‘Collective Responsibilities of Random Collections: Plural Self-Awareness among Strangers’, *Journal of Social Philosophy*, 49: 91–105.⏎

Schweikard, D. & Pettit, P. 2006. ‘Joint Action’, *Philosophy of the Social Sciences*, 36(1): 18–48.⏎

Schwenkenbecher, A. 2014. ‘Joint Moral Duties’, *Midwest Studies in Philosophy*, XXXVIII: 58–74.⏎

Tännsjö, T. 2007. ‘The Myth of Innocence: On Collective Responsibility and Collective Punishment’, *Philosophical Papers*, 36(2): 295–314.⏎

Wringe, B. 2010. ‘Global Obligations and the Agency Objection’, *Ratio*, 23: 217–231.

OceanofPDF.com

7 Legitimate Firms

Stephanie Collins

DOI: 10.4324/9781003612827-8

1 Introduction

When philosophers theorise collective agency, they usually paint on a large canvas. They theorize an incredibly wide range of entities: from nation-states, to firms, to clubs, to universities, and so on. The presumption of philosophers is that we can develop theories of collective agency—and, by extension, collective ethics, obligation, responsibility, virtue, and vice—that will apply to a wide range of different kinds of social entities.[1]

Yet collective agents vary wildly from one another. To take two prominent examples, consider states and business-corporations. While many social ontologists theorize these in the same breath—and while the law takes both to be 'corporate entities' capable of having duties and doing wrong in their own right—these two entity types are clearly disanalogous along many dimensions. For one thing, they have different functions or purposes: states enforce territorial sovereignty, while businesses operate for a profit. Such differences raise the question of whether our theories of collective agency and collective ethics can truly theorize about both entity types at the same time. The underlying issue is: can all the ethical concepts that apply to one type of collective agent (such as states) also apply to other types (such as businesses)?

This chapter explores this question with reference to one specific concept: *legitimacy*. Legitimacy is usually viewed as the sole purview of nation-states. However, philosophers have recently explored the ways in which businesses can fail to meet standards that are sometimes associated with legitimacy—such as through being problematically authoritarian (Anderson 2017) or being insufficiently democratic (Ferreras 2017).[2] Social

media companies, in particular, are said to be confronting a 'crisis of legitimacy' as a result of their role in democratic polarization, privacy violations, and other harmful outcomes (Abiri and Guidi 2023). While these entities do wield power, it is not clear whether they have the intent to govern in the way that states do.

Is it, therefore, a category error if we apply the concept of legitimacy to collective agents other than states—such as, for example, business-corporations? I argue that it is not: we can mold the concept of 'legitimacy' so that it extends to non-state collective agents, including business-corporations. Doing so provides useful conceptual resources for critiquing and guiding the conduct of non-state collective agents, specifically business-corporations. However, our application of the concept of 'legitimacy' to any given collective agent requires that we clarify the *proper function* of that (type of) collective agent. Therefore, this chapter's analysis suggests a more general lesson for social ethics: when extending an ethical concept from one type of entity to another, it will often pay to attend to the different functions served by those different entities. While many concepts —such as 'legitimacy'—might be applicable across a range of entities, this application must proceed with sensitivity to the particularities of the entities involved.

The chapter is structured as follows. First, I outline a theory of collective agency that encompasses both states and business-corporations, drawing on my previous work. This sets the scene for the kinds of socially-constructed entities I will go on to examine: states and business-corporations. Second, I turn to legitimacy. I draw on the recent literature on political legitimacy to develop a broad and ecumenical account of legitimacy. This account will make crucial reference to the *function* of the entity to which we are applying the concept of (il)legitimacy. Third, I apply my account of (il)legitimacy to business-corporations specifically, arguing that a business-corporation is 'legitimate' just in case it adequately performs its function of producing certain characteristic goods via the distinctively corporate structure. This allows us to extend the concept of legitimacy beyond the state, while not conflating legitimacy with justice or with moral perfection. Fourth, I spell out what 'adequately performs' amounts to: I suggest that, at a minimum, a legitimate business-corporation must *respect basic rights* and be *responsive to the needs of stakeholders*, within the domain of the goods

it characteristically produces. This minimal account allows us to distinguish a legitimate business from a perfectly 'just' or 'good' business. I close by discussing how we should treat illegitimate business-corporations.

2 Collective Agency: States and Business-Corporations

Collective agency is a broad phenomenon, arguably encompassing everything from two individuals walking together (Gilbert 1989) or painting a house (Bratman 1992), right up to the complex machinations of intergovernmental organizations. I take a 'collective agent' to be an entity constituted by two or more humans, who are united under a collective decision-making procedure (Collins 2019). In this broad sense, collective agency is ubiquitous. Yet when we are considering the concept of 'legitimacy', we are usually interested in *organizations*, which can be understood as a subset of collective agents. I will take an 'organization' to be a collective agent that involves a large number of people who realize a structure that coordinates divided labour via rules and hierarchical command relations, guided by a collective decision-making procedure (Collins 2023). Organizations are the collective agents that wield deep agential power over our lives—be they public or private organizations—and about which we might have concerns related to the legitimacy of that power.

I will assume realism about organizations: that is, I will assume organizations 'genuinely' exist and should not be eliminated from our ontology by being reduced to a bundle of individual humans and their relations. Broadly, I take realism to be supported by three considerations (Collins 2023: Ch. 1). First, organizations are realizations of structures, where structures are understood as networks of roles in relations, such as might be laid out in a corporation's role chart or a state's constitution. The structure of any given organization is *multiply realizable*, meaning that the role-occupants can be replaced by role-occupants of different types, even while the organization remains identical. For example, a state's government can change even while the state remains the same state. Because the organization survives radical changes in role-occupants, the organization

should not be reduced to a bundle of role-occupants. Second, organizations have inward and outward *explanatory power*: events, facts, and other phenomena, both inside and outside the organization, change in ways that are best explained by references to facts about the organization's structure and/or decision-making procedure, not just facts about individual members. Again, this gives us reason to include the organization in our ontology. Third and finally, organizations satisfy prominent theories of agency—including functionalist, interpretivist, and representationalist theories (Hess 2014). If anything is real, agents are. So, if organizations are agents, then they are real, too.

Because this chapter seeks to take a concept that usually applies to states and extend that concept to business-corporations, it will help to begin by briefly explaining how I understand 'states' and 'business-corporations'. There are two popular definitions of states. First, the Montevideo Convention on the Rights and Duties of States (1933) describes states as having '(a) a permanent population; (b) a defined territory; (c) government; and (d) capacity to enter into relations with other states'. Second, Max Weber's influential definition characterizes a state as 'a human community which (successfully) lays claim to the monopoly of legitimate physical violence within a certain territory'.[3] These definitions are compatible, so I'll assume that a state is something that satisfies both definitions.

Business-corporations are a species of corporation. I'll assume a 'corporation' is any entity that has a legal personality distinct from the humans that underpin it, such that the entity can own property and make contracts in its own right. A *business*-corporation, I will assume, is distinct in having shareholders, who bring in train five important features.[4] First, shareholders invest capital in the corporation (as opposed to investing labour, as in cooperative corporations). Second, this investment gives shareholders the option of ongoing financial rewards (dividends) and the ability to elect board members who oversee management of the corporation. Third, shareholders cannot withdraw their capital ('asset lock-in'), though they can sell their shares: this enables the corporation to engage in long-term planning. Fourth, creditors of the shareholders have no claim to the assets of the corporation ('entity-shielding'). Fifth and conversely, creditors of the corporation have no claim to the assets of shareholders ('limited

liability'). These five features distinguish business-corporations from corporations that are not businesses (such as churches, universities, and so on), and from businesses that are not corporations (such as partnerships and sole proprietorships). From now on, I will use 'corporation' to refer to business-corporations specifically.[5]

I will focus on large and publicly traded business-corporations. It's these entities—including such behemoths as Amazon, Meta, or Google—whose legitimacy is often debated or questioned, both in academic writing and in public debate. Large and powerful businesses are almost all publicly traded corporations, or at least corporations of some form. This is unsurprising: the corporate structure enables huge pooling of capital, which creates the conditions for the wielding of power. And publicly traded corporations raise distinctive legitimacy concerns because the fiduciary duty of managers towards shareholders is in potential tension with the corporation's legitimacy vis-à-vis workers and wider stakeholders. Yet their acts and omissions shape our social, economic, and political order. Thus, the question arises of whether they are legitimate.

3 Legitimacy

The concept of legitimacy is contested by political and legal theorists. As mentioned above, legitimacy is often understood as a 'right to command' or 'right to rule'—where commanding and ruling can be understood as involving the intent to govern within a certain territory. A claim to legitimacy is a claim to be recognized as legitimate by those within the territory over which the right to command is being claimed (Friedman 1990: 64).

If legitimacy were necessarily tied to territory in this way, then it would be a poor fit for the powerful corporations mentioned above. Such a conception of legitimacy encourages us to look away from those to which we haven't formally assigned (or who do not formally assert) a 'right to rule' over a territory, but that nonetheless influence our lives in pervasive ways. To extend the concept of legitimacy to corporations, we need to look elsewhere.

Sometimes legitimacy is tied to *authority*, which might be understood as the ability to give second-order reasons.[6] On this conception, the concept of legitimacy applies only to entities that purport to give second-order reasons. However, again, many business-corporations whose legitimacy is frequently discussed or questioned do not purport to provide second-order reasons—certainly not to outsiders such as customers or other stakeholders, who might reasonably question the legitimacy of the corporation's role in their lives.

In recent decades, work in political theory has sought to look beyond the state, to examine the legitimacy of institutions such as the European Union, international courts, or international human rights institutions (e.g. Føllesdal 2006; Tasioulas 2010; Christiano 2010; Buchanan 2010; Thomas 2014; Føllesdal, Schaffer, and Ulfstein 2014). However, this work, too, has its limitations. Scholars often either take familiar state-based conceptions of legitimacy as their starting point (modifying these conceptions as little as possible to apply to non-state institutions), or else develop a notion of legitimacy that is specific to a particular non-state institution that is the subject at hand.[7] In this way, this work is also not quite as general as we might hope.

One might instead contend that a legitimate business-corporation is simply a business-corporation that is legal by the lights of a legitimate state. On this account, we could meaningfully assess corporations' legitimacy, but it would be reducible to the legitimacy of the state within which it operates. However, many business-corporations are multinational: it is unclear which state or states we should turn to when assessing the legitimacy of multinationals. Moreover, I assume that legitimate states can sometimes get it wrong, either by failing to legitimize morally legitimate corporations or by legitimizing morally illegitimate corporations. This is largely what animates scholarly accusations of corporation illegitimacy (e.g., Anderson 2017; Ferreras 2017). We should not always defer to the state in this arena.

Thankfully, there is a different—more general—conception of legitimacy that has recently been developed in political theory. This conception, I will suggest, can be developed to apply to business-corporations. On this conception, legitimacy theorizing does not start with the state and seek to mechanically apply the resulting theories to non-state

entities. Nor does it theorize different entity types in isolation, as if there were little in common about the legitimacy standards that we might seek to apply across different socially constructed entities. Instead, this more recent work starts at a higher level of generality, aiming to develop a conception of legitimacy that can apply to a range of entities or institutions.

Allen Buchanan is a central figure here, arguing in *The Heart of Human Rights* that the distinctive role of legitimacy—distinct, that is, from justice, fairness, and so on—is that legitimacy solves a 'metacoordination problem' (Buchanan 2013: esp. Ch. 5). The problem arises as follows. We face a first-order coordination problem of how to distribute resources. First-order institutions (such as laws, norms, conventions, practices) allow us to coordinate in order to distribute resources. However, we disagree about which first-order institutions we should have, because we disagree about the resources that should be distributed, the principles that should guide their distribution, and which norms and conventions will best embody those principles. The metacoordination problem is the problem of which institutions we should use to solve our first-order coordination problem.

In Buchanan's framework, legitimacy is the concept we use to pick out those institutions that should be permitted to solve our first-order coordination problem. Legitimacy, therefore, solves our metacoordination problem. Importantly, an institution can be legitimate—it can be permissible for that institution to function to solve our first-order coordination problem—even if that institution is not perfectly just or fair. The point of legitimacy is to enable us to coordinate upon some coordination mechanisms even in the face of robust disagreement, so perfect justice or fairness would be too high a bar. If an institution is legitimate, then it should not be interfered with in its propensity to enable our coordination of distributions of resources.

This conception of legitimacy abstracts away from the state and applies to various other social entities and organizations, especially international institutions. However, the conception doesn't quite capture concerns about business-corporations' legitimacy. For business-corporations arguably do not function to solve our first-order coordination problems about how to *distribute resources*. True, we can think of Google as enabling us to coordinate in the distribution of information; we can think of Amazon as enabling us to coordinate in the distribution of consumer products; and we

can think of Meta as enabling us to coordinate in the distribution of social connections. But not all business-corporations function to enable their customers or users to coordinate. Walmart or Microsoft, for example, do not enable customers or other users to coordinate with one another. Instead, these entities themselves coordinate with customers. Yet the question of these entities' legitimacy can still arise.

For this reason, I prefer to follow the approach developed by Nate Adams, who suggests that 'legitimacy answers a very specific fundamental question: *must we allow this institution to carry on, or may we coercively interfere with it?*' (Adams 2018: 88–89).[8] This way of formulating the question of legitimacy does not wed us to the idea that the candidate for legitimacy must be an institution that solves our first-order coordination problems. Instead, *any* social entity can be subject to the legitimacy question. According to Adams, 'legitimate institutions have the right to function without coercive interference and institutions function via institutional insiders taking up their roles and acting on institutional norms' (2018: 90). This conception of legitimacy is readily applicable to the business-corporations that are my focus.

Now, it's notable that Adams applies his question of legitimacy only to *institutions*, not to *organizations*, which is the category within which I have placed business-corporations. There are many ways of understanding institutions, perhaps the most common of which is that institutions are coordination devices—including devices such as laws, norms, conventions, and practices. If this is what's meant by 'institution', then Buchanan's first-order coordination problem seems to be baked into the idea of legitimacy and therefore baked into any candidate for legitimacy. However, we needn't buy into this restrictive idea of what an institution is. There is a good sense in which any collective agent—and especially any organization—is an institution, namely that it is a structure that enables people to relate to one another.[9] Large public-funded business-corporations like Amazon, Meta, and Google are institutions in this sense, and Adams's question of legitimacy can apply to them.

Different institutions are subject to different legitimacy standards: what it takes for Amazon to be legitimate might be different from what it takes for India to be legitimate, which might differ again from what it takes for

the UN's Human Rights Committee to be legitimate. In later work, Adams emphasizes this. He notes that when we assess an institution's legitimacy, we are assessing it against the standards that are internal to institutions *of that kind*. Adams designates an entity's kind with '*K*'. He writes: 'Legitimacy discourses help us identify bad *K*s and coordinate our response to them. … Illegitimate wins are wins that shouldn't be counted as wins. Illegitimate states are states that shouldn't rule. Illegitimate children are children that shouldn't inherit' (2022: 389). The idea, then, is that the question of whether an entity E is legitimate can be answered only after we have clarified by what standards—i.e., the standards of which *K* (e.g. wins, states, or children)—we are assessing E *and* after we know what the function of *K*s is (e.g. counting as a win, ruling, inheriting).

Building on Adams, I propose the following general account of an object's legitimacy:

> *General legitimacy:* Asking after an object O of kind K's legitimacy-qua-K is asking whether we (prudentially/morally/conceptually) should permit/enable/support O to function in our practices in the way Ks characteristically function, presuming O's internal features remain more-or-less as they are.

If object O is an illegitimate instance of its kind K, then it becomes (prudentially, or morally, or conceptually) permissible to actively block O's functioning in our practices in the way Ks characteristically function, presuming O goes on more or less as it does.

This is a maximally general account of legitimacy, applying to entities well beyond organizations. Consider, for example, that the above account allows us to make sense of the notion of a 'legitimate friend'. If someone is a legitimate friend, then we should (either morally, or prudentially, or conceptually) allow that person to function in our practices in the way friends characteristically function. Friends characteristically function by being trusted, shared with, supported, and so on. So, a legitimate friend is an entity that should be allowed to function in these ways. Notice as well that someone can be a morally, or a prudentially, or a conceptually

illegitimate friend. If the friend is morally legitimate, then the standards of morality deem it permissible for them to function as a friend. If a friend is prudentially legitimate, then the standards of prudence or self-interest deem it permissible for them to function as a friend. And if the friend is conceptually legitimate, then the concept 'friend' should be applied to that person in discourse—whereas, if they are a conceptually illegitimate friend, then they are not a 'real' friend at all.

When we are discussing the legitimacy of socially constructed entities such as states or corporations, we are generally not concerned with prudential or conceptual legitimacy. We are rather concerned with moral legitimacy. Moral legitimacy can be generalized as follows:

> *Moral legitimacy:* Asking after an object O of kind K's moral legitimacy-qua-K is asking whether we morally should permit/enable/support O to function in our practices in the way Ks characteristically function, presuming O's internal features remain more-or-less as they are.

This account is maximally general: it allows us to apply the concept of legitimacy to anything with a function—including tables and chairs. I will speak as if that's the case. For those who find this objectionable, one could limit discussion of 'legitimacy' to only those objects that fall into a kind that *wields great power*. This would still allow the above account to apply to (powerful) corporations. Whether we limit the definition to powerful objects or not, in order to apply this account of moral legitimacy to corporations in a meaningful and informative way, we need to know what corporations' characteristic function is. I turn to this issue next.

4 Functions

The application of the above account of moral legitimacy to business-corporations is not straightforward, because it requires us to be clear on the function of a business-corporation. After all, a morally legitimate business-corporation is one that is morally permitted to function in our practices *as business-corporations characteristically function*. What does this italicized

clause mean? Without answering that question, we cannot know what legitimate business-corporations have a right to do. Without knowing that, we cannot begin to assess the substantive requirements they have to meet to be legitimate.

In answering the question of business-corporations' function, it helps to distinguish between various accounts of 'function' that are prominent in areas such as philosophy of biology or philosophy of mind.[10] Some theories understand a thing's function in terms of the mechanisms that cause the thing to be *reproduced*. For example, on this view the function of a heart is to pump blood because pumping blood is the mechanism that causes hearts to be reproduced. Other theories understand a thing's function in terms of the role the thing plays in some *system*, regardless of whether the thing is likely to be reproduced. Or we might identify a thing's function with the *intentions* that were in the mind of the creator of the thing (obviously, this only works for things that have creators). Finally, we can understand a thing's function in terms of the *values* that are served or promoted by that thing.

I will focus on functions in the last sense described above: a thing's function is the values the thing serves. This is because I am concerned specifically with *moral* legitimacy. If the other three theories of function were followed, that would have problematic implications when theorizing a corporation's legitimacy. For example, a corporation might be *reproduced* because it serves the interests of society's most powerful. But that cannot render the corporation morally legitimate. And the role the corporation plays in our economic system might be to maximize profits for shareholders. Again, there is nothing distinctively moral about this function. And on the intentions-based theory of functions, a corporation's legitimacy is beholden to the intentions of the corporation's creator. But a corporation's creator might have had all sorts of nefarious intentions, the satisfaction of which does not render a corporation morally legitimate.

When asking about how business-corporations characteristically function, then, we are asking about the values they serve. The topic is often approached under the heading of the 'theory of the firm'. Theories of the firm seek to encompass not just business-corporations, but all business types—including partnerships and sole traders. Still, these theories may tell

us something about the function of business-corporations. Transaction cost theory says that firms function to *make transactions efficient* between producers and consumers, via the pooling of labour and resources (Coase 1937). Agency theory says that the firms function to allow capitalists (shareholders) to *invest without high risk*, which is made possible in corporations due to limited liability and the fiduciary duty of managers to pursue the corporation's interests (Jensen and Meckling 1976). Concession theory says that corporations are (or, at least, were originally, when the corporate charter was first invented) created to *develop large projects*—such as railroads or trade (Ciepley 2013). Market failure theory says that firms function to *resolve market failures* (by pooling resources), so have obligations not to exploit market failures (which might imply that firms become illegitimate when they exploit market failures) (Heath 2014). Finally, productivity theory says that corporations function as sites of *norm-governed productivity* and that, as such, they should avoid profiting from or exacerbating their ability to affect politics at the expense of democratic equality (Singer 2018)—which might imply that a firm becomes illegitimate when it profits from or exacerbates its own ability to non-democratically affect politics.

Not all of these theories are put forward as 'theories of the firm'. Heath's and Singer's theories, in particular, are more like theories of business ethics. But all of these theories have something to say about which are the values that firms characteristically serve (namely, the italicized phrases), which might be taken to be a business-corporation's function in the sense outlined above.

However, when we query the legitimacy of Amazon or Google, we don't seem always to be querying whether it should be permitted to function in the ways italicized above. Sometimes we are objecting to something *broader*, such as the effects these corporations have on employees' or consumers' quality of life. These effects can be highly negative—and so, can give rise to complaints of illegitimacy—even while the corporation is doing perfectly well at serving all the values that are italicized above. So, our conception of corporate legitimacy had better make space for these *broader critiques* in some way. Conversely, sometimes when we frame a corporation as illegitimate, we are objecting to something more *specific*. Sometimes, we are objecting to the corporation's stepping outside the

bounds of its usual products or services—as when an arts and crafts store takes a stand on reproductive rights.[11] So, our conception of corporate legitimacy had also better make space for these more *specific critiques*, too.

With both the broad and specific critiques in mind, I propose to understand a given business-corporation's function as *the production of certain characteristic goods or services, via distinctively corporate mechanisms and structures*. The characteristic goods or services will differ from corporation to corporation. This specificity enables us to capture the more specific objections mentioned in the previous paragraph. For example, the characteristic good or service provided by Google is likely to be the distribution of information, while the characteristic good or service provided by Shell is likely to be energy production. Under this proposal, corporations that produce different characteristic goods or services have different functions. This means different corporations fall into different kinds, and so are subject to different standards of legitimacy.

In this way, the difference between an information-providing corporation and an energy-providing corporation is rather like the difference between a friend and a colleague. Someone might be a legitimate colleague without being a legitimate friend (or vice versa), since these two different kinds serve different functions and so have different legitimacy standards: what it takes to be a morally legitimate friend is different from what it takes to be a morally legitimate colleague. Likewise, a corporation might be a legitimate information-providing corporation without being a legitimate energy-providing corporation, because these are two different kinds of corporation that correspondingly have two different sets of legitimacy standards. This enables us to capture the kinds of legitimacy-based criticisms that are made when a corporation goes 'off-script' and engages in an area that is arguably not part of its 'core business'.

In applying this conception of corporate legitimacy, we face the issue of differentiating or specifying the characteristic goods or services produced by a given corporation. This, I think, is likely to be the heart of debates about any given corporation's legitimacy. In assessing whether Google or Amazon or Meta is legitimate, what we're really asking is: what is this corporation *for*? What values does it characteristically serve? What do we want it to do for us? These questions, I suggest, are exactly the questions

that are behind debates about legitimacy. By focusing on these questions, we can see how the question of corporations' legitimacy is perfectly analogous to questions of states' legitimacy. After all, when we question a given state's legitimacy, the questions at issue are the same: what is this state *for*? What values does it characteristically serve? What do we want it to do for us? Until these questions have been answered, we cannot answer the question of what it takes for a state to be legitimate; still less can we answer the question of whether a given state is legitimate. My suggestion, simply, is that the same point holds for corporations. The concept of legitimacy, therefore, can neatly be tracked over from states to corporations.

In identifying a corporation's characteristic goods or services, there is no reason for philosophers to be anything other than maximally pluralistic: we can allow that each actual corporation falls into many vague, concentric, overlapping, and changeable kinds, where each kind is picked out by a characteristic good or service. This allows us to be pluralistic in the kinds of legitimacy questions we ask of any given corporation: we can ask whether Meta is legitimate qua provider of algorithmically determined content, or whether it is legitimate qua facilitator of social networking, and so on. It may well be that many corporations are legitimate when viewed under kind K_1 but are illegitimate when viewed under kind K_2. This enables us to be more targeted in identifying the nature of our normative criticisms.

I've characterized a given business-corporation's function as the production of certain characteristic goods or services, *via distinctively corporate mechanisms and structures*. The distinctively corporate mechanisms and structures were outlined when I characterized business-corporations. Why include these within our definition of a corporation's function, and therefore give these mechanisms and structures a role in defining the kinds by which we assess corporations' legitimacy? Because this enables us to distinguish corporations from charities or state agencies that produce the same characteristic goods or services, but that might be subject to different legitimacy standards. For example, I want to leave open the possibility that charities are subject to legitimacy standards that include, for example, requirements of transparency or requirements that the entity is committed to continue producing the good or service under conditions of adversity. Private organizations might not be subject to quite the same

standards in these regards. If so, then what it takes for (say) an information-producing business-corporation to be legitimate might be different from what it takes for an information-producing charity to be legitimate. This is why a business-corporation's function includes the distinctively corporate mechanisms and structures. (Here I am not firmly committing to the idea that charities and businesses really are subject to different legitimacy standards; I simply want to leave this possibility open.)

If a corporation's function is the production of certain characteristic goods or services (via distinctively corporate mechanisms and structures), then to ask after a given corporation's legitimacy is to ask whether we morally should permit/enable/support that corporation to function in our practices so as to produce those characteristic goods or services, presuming the corporation's internal features remain more-or-less as they are. This has clarified the *question* of corporations' legitimacy. But how are we supposed to *answer* that question? Broadly speaking, we can say that a corporation is morally legitimate just in case it can be expected to *perform its function adequately*, presuming the corporation's internal features remain more-or-less as they are. This notion of 'adequate performance' of the function allows us to accommodate the 'broad critiques' that I mentioned above: roughly, critiques that claim a corporation is illegitimate because of the effects the corporation has on employees' or consumers' quality of life. By contrast, the 'specific critiques' claimed a corporation is illegitimate due to going off-script from its core function. The specific critiques have already been accommodated by the conception of functions outlined in this section, according to which we assess corporations by the specific goods and services they produce. But to make way for the 'broader critiques', we need to spell out the notion of 'adequate performance'.

5 Adequate Performance

A corporation is morally legitimate just in case it can be expected to perform its function adequately, presuming the corporation's internal features remain more-or-less as they are. 'Adequate performance' must be acknowledged as inherently moralized, since, after all, we are dealing with moral legitimacy. Here my suggestion is tentative and cannot be fully

developed in the space I have. Roughly, my suggestion is that adequate performance includes, at a minimum, (1) protecting basic rights within the domain of the corporation's function and (2) procedures that are responsive to the needs of the corporation's stakeholders (internal and external). Let me elaborate on these in turn.

First, legitimacy requires that the corporation protect basic rights, but only within the domain of its function. The restriction to the domain of its function is important because it allows us to distinguish corporations from states. While states arguably must protect basic rights across a broad range of domains in order to count as legitimate, a corporation is not required (or should not be required, at the bar of being deemed 'illegitimate') to protect basic rights across all the domains in which states function. By restricting the rights-protection requirement to the domain of the corporation's function, we stop short of assimilating legitimate corporations to legitimate states. This restriction also allows us to explain why, say, Shell is not rendered illegitimate by failing to protect basic rights in non-energy domains. While many corporations do seek to act philanthropically outside the domain of their function, a failure to do so should not render a corporation illegitimate. Conversely, the requirement that a legitimate corporation must protect basic rights within the domain of its function also means corporations cannot 'buy' legitimacy by acting in other domains. For example, Shell cannot boost its legitimacy-*qua-energy-producer* by (say) funding social projects in disadvantaged communities. Instead, it must protect basic rights *within the domain of its function*, which may often be more costly to its profits than whatever philanthropic ventures it would voluntarily choose to engage in. Legitimacy in one domain cannot be purchased through action in another domain. It also cannot be undermined by the lack of such action.[12]

Second, legitimacy requires procedures that are responsive to the needs of the corporation's internal and external stakeholders. Internally, this will prominently include employees; externally, it will include customers, contractors, suppliers, and wider community members. Again, this requirement prevents the conflation of legitimate corporations with legitimate states, because the net of stakeholders will not include all those whose interests must be considered by the legitimate state. But it will be a

wide net and will include those whose interests are championed by the broad critiques outlined in the previous section.

Importantly, criteria (1) and (2) do not conflate legitimacy with justice, goodness, fairness, or overall moral rightness. Criteria (1) and (2) are much weaker standards, therefore allowing that a legitimate corporation can fail in many other ethical respects. Not only are legitimacy criteria weaker than criteria of justice, goodness, etc.; legitimacy criteria are also distinctively accommodating of pluralism and are distinctively procedural. A commitment to basic rights and procedural responsiveness does not take a stand on controversial conceptions of the good life or the good society, on which legitimacy standards should be neutral. And they do not aim at particular substantive outcomes. This allows legitimacy to play a unique role in our normative assessments of socially constructed entities, including business-corporations.

That said, (1) and (2) are proposed as *minimum* standards for legitimacy. Legitimacy standards above this minimum will be particular to different functionally characterized kinds of corporation, and therefore beyond the scope of the general analysis provided here. Standards above the minimum must also be dynamic across both time and place, since they should be adjustable in light of what's feasible. Again, such feasibility assessments are beyond the scope of a general analysis of corporate legitimacy. However, when the non-minimal legitimacy standards are set for a given corporate type in a given time and place, those non-minimal standards should be such as to accommodate reasonable pluralism about justice, in order that legitimacy can continue to play its distinctive role.

To see the substantive implications of criteria (1) and (2), it might help to compare them with the criteria of legitimacy offered by Allen Buchanan (2018: 59–60). Buchanan's criteria are presented as if they are intended to apply to all institutions. However, he doesn't mention business-corporations specifically, and he seems mainly to have international institutions in mind. Nonetheless, his criteria serve as a useful counterpoint. His criteria are:

(1) the institution provides better benefits than the *non- institutional* alternative;

(2) its performance accords with its *professed* goals;

(3) it doesn't violate basic rights;

(4) it *originated* in an acceptable way;
(5) it has *good procedures* for resolving disputes.

In my view, these criteria would hold corporations to too high a bar in some respects, and too low a bar in others. Let me explain these in turn.

First, that a corporation provides benefits better than a non-institutional (non-organizational) alternative is a very low bar. When assessing an entity's legitimacy, we should not compare it with the state of nature, but rather its feasible alternatives. This is why I mentioned feasible alternatives when noting the non-minimal legitimacy standards that will apply to corporations on a kind-by-kind basis.

Second, Buchanan's reference to an entity's 'professed' goals again is too low a bar: as mentioned above when discussing the intentions-based theory of corporations' functions, we should not allow an entity to dictate its own function, goals, or values; to allow corporations to do this would be to allow them to manipulate moral requirements to suit their professed goals.

Third, Buchanan's requirement that an entity not violate basic rights accords with my criterion (1), while again being a lower bar. I require not just that an entity doesn't negatively *violate* basic rights, but that it positively *protects* basic rights within its domain. After all, corporations must provide some positive goods in order to be allowed to exist in our world; unlike humans, they do not have any default right to exist.

Fourth, Buchanan's requirement about the entity's origins holds it to too high a bar. Of course, a corporation with morally wrongful origins can have strong reparative obligations in light of that. Here, we might think of businesses whose origins were mired in colonialism or white supremacy. Such wrongful origins need not undermine a corporation's right to function today, *as long as* the relevant reparative actions are taken. It's essential that reparative actions are taken in order for the corporation to gain legitimacy. But if those actions really are meaningfully taken, then the fact of unacceptable origins need not be a death knell for legitimacy.

Finally, Buchanan requires that a corporation has good procedures for resolving disputes. Here again I would say the bar is set too low: the procedures should not just concern resolving disputes, but more generally being responsive to stakeholders' needs. Of course, Buchanan is imagining

international institutions resolving disputes between states. There is no question of international institutions needing to be responsive to states' needs because they are under the control of states, so are necessarily responsive to states' needs (although not responsive equally to the needs of all states). This is presumably why Buchanan speaks only of resolving disputes. But this won't do for corporations, which have extensive power over people who do not control them.

6 Conclusion: The Implications of Illegitimacy

To sum up, I have proposed that a business-corporation is legitimate just in case it can be expected to perform its function adequately, presuming the corporation's internal features remain more-or-less as they are. Its function is the production of certain characteristic goods or services, via distinctively corporate mechanisms and structures. Adequate performance of this function includes—as a minimum standard—(1) protecting basic rights within the domain of the corporation's function and (2) procedures that are responsive to the needs of the corporation's stakeholders (internal and external).

What follows for corporations that are deemed illegitimate? Returning to Adams's account, illegitimate institutions '(1) lack rights to function and to pursue their institutional goals, so (2) others correlatively lack a duty of non-interference, meaning that (3) it is permissible to coercively interfere with these institutions and their members, thereby preventing the members from coordinating their action by following institutional norms, ultimately resulting in the dismantling of the institutions such that they no longer exist' (2018: 92–93). In the case of corporations, individuals' potential responses to illegitimacy look quite different—and potentially more potent—than individuals' potential responses to illegitimate states (let alone illegitimate international institutions).

For parties external to the corporation, the fact of corporate illegitimacy licenses such actions as boycotts, protests, petitions, sabotage, naming and shaming, and (in the case of states that see a corporation is illegitimate) regulation. In general, the fact of corporate illegitimacy licenses the removal of the social license to operate. For parties internal to the

corporation, the fact of corporate illegitimacy licenses such actions as whistleblowing, refusal to follow orders, and sabotage. It also licenses attempts from the inside to overturn the presumption that the internal features will remain more or less as they are: an illegitimate corporation can be rendered legitimate by having its procedures, structures, and norms revised in such a way that it can be expected to adequately perform its function in the future. The categories of legitimacy and illegitimacy have practical import when applied to business-corporations.[13]

Notes

1. See, e.g., French 1984; List and Pettit 2011; Isaacs 2011; Tuomela 2013; Tollefsen 2015; Hess 2018; Hindriks 2018; Collins 2023. For example, in their hugely influential contribution, List and Pettit (2011) take their theory of group agency to apply to: states; 'trade unions, professional associations, and industry groupings'; 'firms and corporations'; 'voluntary associations …; recreational bodies …; and more formal bodies like schools, churches, and universities' (2011: 40). They analyze none of these in-depth. Other authors are less generalizing; for example, Hindriks (2018) argues that not all collective agents are *moral* agents, since (he argues) a socially constructed entity has moral agency only if it has policies that systematically attend to normative considerations. But Hindriks does not apply his criterion for moral agency to the differences between, say, states, firms, and universities. Other theorists of collective agency—such as French (1984) and Hess (2014)—focus their discussion on firms, yet provide criteria for moral agency that are applicable to socially constructed agents other than firms.⏎
2. Democracy is perhaps where the state-firm analogy has been most thoroughly explored and questioned, with several authors

debating whether the power of firms is relevantly analogous to the power of states, such that firms (like states) face a moral requirement to be democratic. (See, for example, Dahl 1985; Landemore and Ferreras 2016; Ferreras 2017; Frega, Herzog, and Neuhauser 2019; González-Ricoy 2022; Cockburn and Preminger 2023) While democracy and legitimacy are related—perhaps an entity is legitimate only if it is democratic—I do not assume that legitimacy presupposes democracy. This enables me to theorize the state-firm analogy in legitimacy, independently of whether the state-firm analogy in democracy also succeeds.⏎

3. Weber 1919: ¶4.⏎
4. Obviously, the details differ between jurisdictions. Here I follow Claassen (2022).⏎
5. I have not mentioned *the profit motive* when characterizing business-corporations, because it's debatable whether business-corporations are defined by the profit motive (Blair and Stout 1999). I'll return to this issue below, when discussing the function or purpose of business-corporations.⏎
6. Razian second-order reasons are reasons 'to act for a reason or to refrain from acting for a reason' (1999: 39).⏎
7. This observation is made by Adams (2018).⏎
8. Adams (2018) follows Buchanan (2013) very closely, but it's worth noticing that Adams' formulation of the question, as just quoted, does not rely on Buchanan's idea that a legitimate entity is one that solves our first-order *coordination problem*. See similarly Adams (2020).⏎
9. Indeed, Adams at one point analogizes institutions to individual humans (2018: 90)—an analogy that is congenial to those who view organizations as collective agents that are analogous to individual agents. And he names 'a private social club, a local grocery co-op, the temple on the corner, and so

on’ as examples of institutions (2018: 92). So, I believe he means for his account of institutions’ legitimacy to apply to organizations as I have characterized them.⏎
10. The following taxonomy of functions comes from Silver (MS).⏎
11. Here I’m referring to the famous case of *Burwell v. Hobby Lobby Stores, Inc*. 573 U.S. 682 (2014). For objections to the US Supreme Court’s decision in that case, see for example Strine 2015.⏎
12. I assume ‘protecting’ basic rights is more demanding than merely ‘respecting’ them (which all agents must do for all rights across all domains), but less demanding than ‘promoting’ them (which is a more aspirational, never-ending, and demanding task). ‘Protecting’ requires putting some safeguards in place to ensure other agents respect the rights.⏎
13. For helpful comments on drafts of this chapter, I thank audiences at the Australian National University, Groningen University, and the Legal Philosophy Trident.⏎

References

Abiri, Gilad and Sebasitan Guidi. 2023. ‘From Network to a Dilemma: The Legitimacy of Social Media’, *Stamford Technology Law Review*, 26(1): 92–142.⏎

Adams, N.P. 2018. ‘Institutional Legitimacy’, *Journal of Political Philosophy*, 26(1): 84–102.⏎

Adams, N.P. 2020. ‘Legitimacy and Institutional Purpose’, *Critical Review of International Social and Political Philosophy*, 23(3): 292–310.⏎

Adams, N.P. 2022. ‘The Concept of Legitimacy’, *Canadian Journal of Philosophy*, 52(4): 381–395.⏎

Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (and Why We Don’t Talk about It)*. Princeton: Princeton University Press.⏎

Blair, Margaret M. and Lynn A. Stout. 1999. ‘A Team Production Theory of Corporate Law’, *Virginia Law Review Association*, 85(2): 247–328.⏎

Bratman, Michael. 1992. ‘Shared Cooperative Activity’, *Philosophical Review*, 101(2): 327–341.⏎

Buchanan, Allen. 2013. *The Heart of Human Rights*. Oxford: Oxford University Press.⏎

Buchanan, Allen. 2018. ‘Institutional Legitimacy’, *Oxford Studies in Political Philosophy*, 4: 53–78.⏎

Buchanan, Allen. 2010. ‘The Legitimacy of International Law’, in Samantha Besson and John Tasioulas (eds), *The Philosophy of International Law*, pp. 79–96. Oxford: Oxford University Press.⏎

Christiano, Thomas. 2010. ‘Democratic Legitimacy and International Institutions’, in Samantha Besson and John Tasioulas (eds), *The Philosophy of International Law*, pp. 119–137. Oxford: Oxford University Press.⏎

Ciepley, David. 2013. ‘Beyond Public and Private: Toward a Political Theory of the Corporation’, *American Political Science Review*, 107(1): 139–158.⏎

Claassen, Rutger. 2022. ‘Political Theories of the Business Corporation’, *Philosophy Compass*, (1): e12892.⏎

Coase, Ronald. 1937. ‘The Nature of the Firm’, *Economica*, 4(16): 386–405.⏎
Cockburn, Patrick J.L. and Jonathan Preminger. 2023. ‘Migration and Demos in the Democratic Firm: An Extension of the State-Firm Analogy’, *Political Theory*, 51(3): 557–580.⏎
Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.⏎
Collins, Stephanie. 2023. *Organizations as Wrongdoers: from Ontology to Morality*. Oxford: Oxford University Press.⏎
Dahl, Robert H. 1985. *A Preface to Economic Democracy*. Berkeley, CA: University of California Press.⏎
Ferreras, Isabelle. 2017. *Firms as Political Entities: Saving Democracy through Economic Bicameralism*. Cambridge: Cambridge University Press.⏎
Føllesdal, Andreas. 2006. ‘The Legitimacy Deficits of the European Union’, *Journal of Political Philosophy*, 14: 441–468.⏎
Føllesdal, Andreas, Johan Karlsson Schaffer, and Geir Ulfstein (eds). 2014. *The Legitimacy of International Human Rights Regimes*. Cambridge: Cambridge University Press.⏎
Frega, Roberto, Lisa Herzog, and Christian Neuhauser. 2019. ‘Workplace Democracy: The Recent Debate’, *Philosophy Compass*, 14(4): e12574.⏎
French, Peter. 1984. *Collective and Corporate Responsibility*. New York: Columbia University Press.⏎

Friedman, Richard B. 1990. ‘On the Concept of Authority in Political Philosophy’, in Joseph Raz (ed.), *Authority*, pp. 56–91. New York: New York University Press.⏎

Gilbert, Margaret. 1989. *On Social Facts*. Princeton, NJ: Princeton University Press.⏎

González-Ricoy, Iñigo. 2022. ‘Little Republics: Authority and the Political Nature of the Firm’, *Philosophy and Public Affairs*, 50(1): 90–120.⏎

Heath, Joseph. 2014. *Morality, Competition, and the Firm: The Market Failures Approach to Business Ethics*. Oxford: Oxford University Press.⏎

Hess, Kendy M. 2018. ‘Does the Machine Need a Ghost? Corporate Agents as Nonconscious Kantian Moral Agents’, *Journal of the American Philosophical Association*, 4(1): 67–86.⏎

Hess, Kendy M. 2014. ‘The Free Will of Corporations (and Other Collectives)’, *Philosophical Studies*, 168: 241–260.⏎

Hindriks, Frank. 2018. ‘Collective Agency: Moral and Amoral’, *Dialectica*, 72(1): 3–23.⏎

Isaacs, Tracy. 2011. *Moral Responsibility in Collective Contexts*. Oxford: Oxford University Press.⏎

Jenson, M.C. and W.H. Meckling. 1976. ‘Theory of the firm: Managerial Behavior, Agency Costs and Ownership Structure’, *Journal of Financial Economics*, 3: 305–360.⏎

Landemore, Helene and Isabelle Ferreras. 2016. ‘In Defense of Workplace Democracy: Towards a Justification of the Firm-State Analogy’, *Political Theory*, 44(1): 53–81.⏎

List, Christian and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.⏎

Singer, Abraham. 2018. *The Form of the Firm: A Political Theory of the Corporation*. Oxford: Oxford University Press.⏎

Strine, Leo E. 2015. 'A Job Is Not a Hobby: The Judicial Revival of Corporate Paternalism and Its Problematic Implications', *Journal of Corporation Law*, 41: 71–115.⏎

Tasioulas, John. 2010. 'The Legitimacy of International Law', in Samantha Besson and John Tasioulas (eds), *The Philosophy of International Law*, pp. 97–116. Oxford: Oxford University Press.⏎

Thomas, Christopher A. 2014. 'The Uses and Abuses of Legitimacy in International Law', *Oxford Journal of Legal Studies*, 34: 729–58.⏎

Tollefsen, Deborah. 2015. *Group as Agents*. Cambridge: Polity.⏎

Tuomela, Raimo. 2013. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press.⏎

Weber, Max. 1919[1946]. 'Politics as a Vocation', in H.H. Gerth and C.W. Mills (eds and trans), *From Max Weber: Essays in Sociology*, pp. 77–128. New York: Oxford University Press.⏎

OceanofPDF.com

8 The Problem of Overlapping Collectives

Luke Roelofs

DOI: 10.4324/9781003612827-9

1 The Problem of Overlapping Collectives

Suppose that a hotly contested election has just happened, and a new president is set to be inaugurated. We often find ourselves deeply invested in understanding how this happened. For a start, we look for causes: why did this happen? Moreover, because this is not simply a natural phenomenon with causes but a social event produced by human agency, we can also ask: who is responsible? Even further, if we see the election of this candidate as, in some sense, a collective action, we can meaningfully ask who the agent is: who did this? This quest for agential responsibility need not immediately constitute a quest for moral responsibility, for someone to *blame* or give *credit* to, but if we are interested in moral evaluation, identifying the agent will be a necessary precondition.

There is no shortage of candidates. Indeed, there seem to be too many candidates, with no principled way to identify the right collective agent.[1] We might say that, since this is a democratic system, the electorate 'has decided to elect this candidate', thereby assigning primary responsibility to a very large group called 'the electorate', or to the smaller group of 'people who actually voted'. Or we might restrict this claim to a smaller subset, like 'people who voted for this candidate', or even to 'people who voted for them, chose not to vote, or voted third-party', i.e. everyone who had the opportunity to contribute to electing the candidate's only feasible opponent, and chose not to.

But it's also very common to describe this event by ascribing agency to the candidate themselves, or their campaign team, or their party. They won: they were engaged in a contest of some sort, metaphorically a race or a battle, and they succeeded in what they were trying to do, namely win it. We might also say this about a somewhat larger group, including not only people officially affiliated with the campaign or party, but also anyone else who worked to help that candidate win. We might include foreign leaders like those we think hoped for this outcome and took various known or unknown measures to contribute to making it happen, whether formal endorsements or choices about when to end or prolong a war (see, e.g., Farrell 2017). They too might count as part of the team that won, i.e. that succeeded in doing what they worked together to do.

We might even assign responsibility to collectives including those whose stated or presumed aims run directly counter to the winning candidate's. Vote-splitting candidates, misguided campaigners, unimpressive opponents, or the 'mainstream media' might be blamed for 'having done this' by virtue of some action they took that is thought to have made it more likely.

My aim in this chapter is to delve into one particular theoretical puzzle raised by this sort of attempt to assign agency for a diffuse, large-scale collective action. Of course, there are many theoretical puzzles they raise, and many difficulties and objections they might face. We might object to some or all such assignments of agency on numerous grounds. Just to name five quickly:

- The actual causal dependencies at stake are often unclear: We just don't know what effects were in fact produced by a given public statement, a given diplomatic phone call, or a given campaigning decision in a swing state.
- Responsibility might be thought to depend on knowledge, and many voters might well have been under various significant factual misapprehensions about the candidates' policies or the state of various national metrics.
- Responsibility might be thought to depend on freedom to choose, and the emergent forced-choice character of many electoral decisions, where voting for one candidate becomes the only way to keep out the

other, might render some voters' or volunteers' decisions less than fully free.

- Responsibility might be thought to depend on intention, and many people whose actions might have in fact helped a certain candidate to win (e.g. opposing politicians who perhaps made errors in campaigning) were intending to do just the opposite.
- Collective responsibility in particular might be thought to depend on coordination and mutual knowledge, and it's often unclear how much of this there was among various actors: a newspaper or foreign government might have tried to help a given candidate win, in expectation of preferential treatment later, without ever explicitly negotiating or coordinating with them.

These considerations could be good reasons to doubt, reject, or qualify many of the above attributions of responsibility for an election result. Some groups might, at most, be said to have unknowingly or unintentionally caused this. But I am interested in a different problem: the sheer diversity of candidate collective agents. Suppose, for instance, that we tighten up our ascriptions of responsibility to require positive intention and active contribution. We still face a seemingly impossible task if we try to specify exactly which collective is responsible. There is the candidate himself, and a core campaign team around him, and then the diffuse set of people who volunteered, funded, voted, or otherwise helped out. There are many individuals whose membership in the relevant collective is debatable – who contributed a little, but perhaps not quite enough, or who contributed but with some unrelated motive or factual misapprehension slightly too dominant in their minds. It seems arbitrary to draw any precise line around those who should be viewed as participants in the collective act of winning the election or electing the candidate, and those who merely made their success more likely. And someone might think that this is a good reason to deny that there was any such collective act: individuals and smaller, formally defined groups might have performed actions, but the election itself is just a causal result of their actions, not a collective action itself. This is the line of thinking I want to resist.

Electing a president is only one example. A similar worry about there being too many candidate collective agents will arise for almost any large-

scale collective result which is brought about by a diffuse or informal group. A boycott, a riot, a social movement, or the founding of a new genre, subfield, or school of thought will all involve both some heavily involved ‘ringleaders’ and also various swathes of more or less loosely involved participants, such that attempts to specify any definite collective agent seem doomed. And even larger phenomena raise the same problem. For example, many people believe that women as a gender are oppressed. Oppression seems to be a collective act, so we might then ask: who oppresses women? Who is the agent of this deeply significant collective act? Is it ‘men’? Or the smaller set of ‘sexist men’? Or sexist people, of all genders? And how sexist does one have to be to qualify as a participant in this act? Or will everybody count, so that the agent of oppression is some all-encompassing whole like ‘society’?

Consider also the idea of ‘social construction’. Various things are said to be social constructs: to exist by virtue of our practices and conventions, and possess whatever degree of reality they possess only because *we* have made them.[2] But who is this ‘we’? How should we specify the collective responsible for constructing a given social construct? As with the above examples, it’s important to recognise that the contributions individuals make are often unintentional, unconscious, or just empirically hard to determine. But even independently of these factors, any attempt to actually specify a definite collective agent would seem faintly absurd, and would again face the problem of multiple equally good candidate specifications.

In light of this profusion of overlapping candidate collectives, it might be tempting to try and think about social construction, or oppression, or even elections, in non-agential terms, as simply something that happens, or a mere consequence of many individual actions, rather than something that ‘we’, or anybody, does. In Section 2, I will argue that this would be too hasty, and potentially a political mistake. In Section 3, I will examine how philosophers have addressed a puzzle about individual people that turns out, somewhat surprisingly, to be closely analogous to this puzzle about diffuse collective actions, and in Sections 4 and 5, I show how to extend that solution to the collective case.

2 Bad Responses to the Problem

Let me try to state this problem, the ‘problem of overlapping collectives’, a bit more precisely. It arises when we have a phenomenon of interest which seems, at least at first glance, to be both (i) collective and (ii) agential, i.e. something which seems to have been actively and intentionally done, but not simply by one individual. But if there is a collective act, there must be some sort of agent – not necessarily a persisting or well-organised one, but some sort of answer to the question ‘who did this?’ The problem of overlapping collectives arises when, for any candidate answer identifying a set of people S_1 (or entities constituted by a set of people, like departments, clubs, parties, battalions, etc.),[3] we can also identify one or more sets of people S_n which meet the following three conditions:

1. S_n is just as good a candidate as S_1 for being the agent of the apparent collective act.
2. S_n is clearly not strictly identical to S_1, e.g. because they have different memberships.
3. It does not seem plausible to identify S_1 and S_n as two separate collectives interacting.

For example, consider the set comprised of all people who voted for a given candidate, and the set comprised of all people who voted for them, plus one foreign leader who tried to help them win. These two sets are both (1) plausible candidate answers to the question ‘who elected this candidate?’, and they seem about equally plausible to me. But they are (2) clearly not identical, since one of them contains a person who is not included in the other. And (3) it would be absurd to say, ‘The president was elected by two distinct groups, their voters and also their voters plus this foreign leader’. That is, this case is not like, say, a kingdom which collapses because there happens to be an uprising and an external invasion at the same time, so that two distinct and separate military forces are equally good candidates for being the agent of its destruction. Intuitively, ‘their voters’ and ‘their voters plus this foreign leader’ are two ways of specifying the boundaries of a single group, two ways of defining who counts as a member of what is intuitively a single collective. The problem is that if both collectives exist and are plausible candidate agents, then we seem to have two agents for the

same action. Indeed, we will likely have many more than two – consider the set ‘all their voters except for Steve, a late-decider in a safe state, who flipped a coin to decide how to vote’. There’s some plausibility in saying that, since Steve made so little difference to the outcome, and cared so little about the outcome, he is not really participating in the collective action. But there are probably many Steves, and if ‘all their voters except Steve’ is our third candidate agent (and ‘all their voters except Steve, plus this one foreign leader’ is our fourth), then ‘all their voters except Stu’ is another, and ‘all their voters except Steve and Stu’ is another, and before long we will have dozens or hundreds of technically distinct collectives. And this seems like a problem: answering the question ‘who did this?’ with a list of hundreds of almost-identical sets of people is not answering it well.

Let me consider and reject two responses to the problem of overlapping collectives before sketching my preferred solution in the next section. The two responses I will reject are:

1. The problem can be solved by drawing a ‘fuzzy boundary’, so that some people have intermediate, or even indeterminate, degrees of membership in the relevant collective.
2. The problem just shows that it’s a mistake to try and understand these phenomena in agential terms – they are better seen as products of impersonal social structures.

The first response I will reject appeals to the idea of a fuzzily bounded set, in which different individuals count as participants to different degrees. After all, people constantly work together in ways that are loosely and informally coordinated, or diffusely and unconsciously coordinated, but still genuinely and socially coordinated.

In line with this, in previous work (Martens and Roelofs 2018) Judith Martens and I have suggested that many social phenomena represent ‘implicit coordination’, a sort of ‘quasi-joint action’ in which what each participant aims for overlaps at least partly with what others aim for (as opposed to the fully shared aim of a fully joint action) and acting with partial dependence on the others’ actions (as opposed to acting only on the condition that others do, as in a fully joint action). The greater the degree of aim-sharing and interdependence, the more fully collective in character is

the action. Larger-scale phenomena might typically be less coordinated and thus less collective, closer to a mere aggregation of many individual actions, but they can exhibit a range of degrees of collectivity.

If we can accept variable degrees of participation among different individuals, perhaps we can side-step the question ‘who is in the collective that did this?’, replacing it with the question ‘who had what degree of participation in doing this?’ In our election example, we might say that the candidate and various campaign managers are heavily involved, diehard voters moderately involved, and Steve the coin-flipper less fully involved. Calculating an individual’s degree of participation will likely not be a very precise matter, but we might think it involves, at the very least, combining some measure of how big their causal contribution is with the two measures noted above for degrees of collectivity: how far their aims overlapped with those of others involved, and how far their contribution encouraged, or was encouraged by, those of others. Probably the best way to reckon this is multiplicative, so that (e.g.) even if some mis-step by an opponent made a large causal contribution to getting a candidate elected, that would not be enough to count the opponent as even moderately involved in the collective action of electing him, since their intention was presumably the complete opposite, while conversely some zealous supporter in a foreign country who did nothing and made no contribution to the result also does not count as involved, for the opposite reason.

However, while this is potentially an illuminating move (and one which I will make use of in my preferred response in the following sections), it does not actually remove the problem of overlapping collectives. It just adds more details and more parameters to work into our answer to the question ‘who did this?’ We can say ‘if you mean to ask who was involved to X degree, then it would be these people… but if you mean to ask who was involved to a certain lower degree, then it would be this larger set of people…’. But this is still an awkwardly long and unhelpful answer – not to mention, given the imprecision involved in calculating degrees of participation, there will potentially be multiple sets of people who are equally good candidates for being the agent, even given a specific degree of participation. To put it another way, it will still be true that there are many different collectives, which will have good claims to being the agent of the action of interest. We may have analyzed their differences more carefully,

and that is useful. But the basic fact of multiplicity remains to be confronted.

The other response I want to reject is to doubt that such diffuse actions, spread over so many people and coordinated by such variable and informal sorts of relations, should be thought of as collective acts in the first place. The paradigmatic sorts of collective act in most philosophical discussions are of two sorts. Either they are joint intentional actions, where each participant is aware of the others, and intends to cooperate with them towards a goal they all recognise as shared, or they are the actions of enduring institutional agents, whose formal structures give them a stable and distinctive group perspective.[4] What we are considering now is neither of those things, and so might seem to be 'collective' only in an inordinately weak sense: Many individual actions whose effects happen to add up to a larger result. Indeed, the problem of overlapping collectives might itself be a good reason to drop the framework of collective agency for thinking about cases like these. Maybe we are just taking a lens that is useful on the individual level and misapplying it to the collective level. Perhaps instead we should take a strictly impersonal view, at least outside of organised institutions. For diffuse actions like oppressing a whole gender, or 'socially constructing' something, the best thing to identify as the agent is something all-encompassing like 'society', or 'culture', or 'the system'.

One reason to reject this response is that versions of the overlapping collectives problem can arise even for small, cohesive groups. Suppose that in the midst of a large heated protest, after one protestor has been arrested, another protestor looks at six others, points, and shouts 'let's get them back!', initiating a coordinated rush at the police line to de-arrest the person arrested. This seems like a clear case of joint action, with a clear and explicit shared intention, and each acting only because the others are. But suppose that of the six people explicitly addressed, only five join in fully. The sixth makes a half-hearted start towards the police, in a way that helps throw them off and might have helped the others break through; another protestor sees what's happening and joins in uninvited, and another joins halfway through, while others flock to the area, to ensure that when the arrested protestor is pulled out, they are swiftly lost in the crowd. Here we have six people who are clearly participants, and many others who are

clearly not participants, but several others who have varying degrees of participation, such that they could equally well be included or not included among the participants of the action. So, who completed the de-arrest? We could give a dozen answers, mentioning different combinations of people, and so once again we face the challenge of figuring out how best to describe the situation.

But a more basic reason to reject this response is that there's no reason why explicit joint actions, or actions by institutionally structured group agents, should exhaust our conception of collective action. There is a meaningful difference between social conditions which people have made and continue to actively sustain and reproduce, and natural phenomena that simply exist or have unintended consequences. In particular, the former invites a range of reactive attitudes – blame, praise, anger, resentment – that would be inappropriate to the latter. And so, if any sufficiently large collective phenomenon without a well-defined institution responsible for it is simply treated as the mechanical product of 'society', we face a sort of 'responsibility gap',[5] or at least a difficulty in knowing where to direct our reactive emotions.

I don't think this is a purely academic problem: the frustration of not knowing who to blame for large-scale, diffuse, but still seemingly deliberate social phenomena is a real psychic burden for many people. Various bad options readily present themselves. If we retain the idea that being part of the agent of injustice is a bad thing, a sort of guilt or complicity, then we might end up feeling that whatever a given person knows or wants or does, they are part of 'society' and thus part of the agent responsible for injustice. But this risks blaming too many people, including the victims of oppression and plenty of people who intuitively should be thought innocent, either because they bear heavy personal costs to fight against oppression or because they are, for instance, three years old. We should not allow opposition to injustice to become a doctrine of original sin.

The opposite mistake is to simply drop blame from the equation entirely, effectively saying nobody is to blame, thereby letting too many people off the hook for their various degrees of complicity, and letting a distinctively moral sort of wrong be naturalised as 'just how things are'. People may react to this sense of a complete evasion of blame by looking for specific

people to hold fully responsible, potentially distorting reality to sharply divide the complicit from the innocent, or insisting that large-scale problems are entirely due to a small and nefarious elite. The anarchist folk singer Utah Phillips is often quoted as saying 'The earth is not *dying*, it is *being killed*, and those who are killing it have names and addresses [emphasis added]'.[6] But this risks overstating the degree of influence exercised by specific individuals, and the degree of intent and coordination they exercise. In the extreme, looking for people who are explicitly and individually responsible for large-scale social harms risks lapsing into conspiratorial thinking.

For me, the above reflections underline the importance of thinking clearly about diffuse collective agency, without either distorting or abandoning its agential character. This chapter aims to make a small contribution to this task by identifying and addressing the problem of overlapping collectives.

3 Manyism and Overlapping Individuals

I think the problem of overlapping collectives is really just one instance of a general issue in metaphysics, often called 'the problem of the many', or more broadly the 'puzzles of multiplicity'.[7] Arguments are advanced which seem to show that any seemingly individual object (including, most worryingly, a human brain or body) can be analyzed into a vast 'cloud' of many overlapping entities of the same sort, differing in their parts but sharing the most important parts. For example, it might be argued that if the whole human being is a conscious subject, then so is the human brain, and the human head, and the top half of a human being, and the whole human being except one hair, and the whole human being except one dead skin cell on the edge of flaking off, and the whole human being plus one dead skin cell that just flaked off, and so on *ad infinitum*. This sort of puzzle arises for any entity which is composite, and especially for composite entities with fuzzy boundaries. (Consequently, the desire to avoid this result is sometimes taken as a reason to accept substance dualism, and hold that the human mind is mereologically simple and sharply bounded, see e.g. Zimmerman 2010; Simon 2017). So, it should in a sense be unsurprising

that it arises for social collectives, which are very evidently composite, and plausibly often very fuzzily bounded.

My approach to these puzzles is heavily indebted to David Lewis's treatment of them. Lewis (1993) considers and rejects several attempts to salvage singleness by collapsing the many somehow into one: by relativising identity and distinctness to a sortal category; allowing many portions of matter to each constitute the same living body; positing a metaphysically vague entity with intrinsically indeterminate boundaries, and so on. I will not dwell on these solutions, because I am convinced by Lewis's reasons for rejecting them, and because I think they are even less attractive for the problem of overlapping collectives, which do not even intuitively *seem* to be metaphysically single and unified.

Instead, I am interested in applying Lewis' preferred solution, which involves the idea of counting not by identity but by 'almost identity', i.e. by high levels of mereological overlap. Lewis starts from the observation that numerical statements like 'there is one cat on the mat' (which the puzzles of multiplicity seem to undermine, by showing that there are actually many cats) are standardly understood logically in terms of identity and distinctness, but could instead be understood in terms of very extensive overlap.

> Assume our cat-candidates are genuine cats. […] No two of them are completely identical. But any two of them are almost completely identical; their differences are negligible, as I said before. We have many cats, each one almost identical to all the rest.
>
> Remember how we translate statements of number into the language of identity and quantification. 'There is one cat on the mat' becomes 'For some x, x is a cat on the mat, and every cat on the mat is identical to x'. That's false, if we take 'identical' to express the complete and strict identity that lies at the end of the spectrum. But the very extensive overlap of the cats does approximate to complete identity. So, what's true

> is that for some x, x is a cat on the mat, and every cat on the mat is almost identical to x. In this way, the statement that there is one cat on the mat is almost true. The cats are many, but almost one. By a blameless approximation, we may say simply that there is one cat on the mat.
>
> (Lewis 1993: 33)

This proposal involves three basic steps. Step 1: Accept the weird result that there are actually many things where we thought there was one. Step 2: Insist that the near-complete overlap, or 'almost identity' between these many things substantially reduces the weirdness of this result. Step 3: Explain how common-sensical things can still be said and come out true despite that weird result.

My proposal, in a nutshell, is to do the same for the problem of overlapping collectives. Step 1 is to admit that the reason it is hard to identify a specific collective agent for diffuse social actions is that there are in fact a great many agents: many collectives who all count as performing that action. There are not just many individual agents making up a single collective agent; the many individual agents make up a far greater number of collective agents, and each of these 'did that'. Step 2, I hope, is less urgent for social collectives than for individual people: it is not so surprising that there are many overlapping social groups. This leaves step 3, of explaining how we can still usefully talk and think about diffuse overlapping collectives. It is to this challenge that I now turn.

4 Many Manies, but Almost One Many

The key requirement for the manyist solution to the problem of overlapping collectives is to explain how we can make quantified statements about, and singular reference to, collective agents (including statements like 'the people responsible' or 'those who did this') based on counting not by strict identity and distinctness but based on some other counting relation.

What is the relevant counting relation? I would suggest that there should be *substantial overlap of members, weighted by degree of participation* (for

brevity I will express this as ‘substantial overlap’). So, if I refer to ‘the people who elected this candidate’, this need not presuppose that there is a unique collective which elected them, as its singular form would suggest. Rather, it presupposes that there is at least one collective which elected the candidate, and which is such that all other collectives which elected them substantially overlap with it. And statements of the form ‘the people who got this guy elected are F’ will then be true if there is such a collective, and if its members are F.

The definition of substantial overlap needs to be clarified in a few ways. First, I take proper parts of something to overlap with it entirely, even if they are small. Thus, the collective ‘people who elected this candidate’ overlaps entirely with the candidate themselves, and with Steve (the swing voter mentioned above), even though these two individuals do not overlap at all with one another.

Second, I take ‘substantial’ to mean something like ‘large enough for present purposes’. Exactly what counts as large enough might depend on context: in casual conversations, the standards for overlap to be substantial might be low, while for works in political sociology they might be much higher. As a result, sweeping statements about ‘the people who elected this candidate’ might count as true in casual conversation but not in political sociology writings, because no collective of which the statement is true overlaps with all other relevant collectives to an extent that is significant by the latter’s standards.

Third, an individual’s degree of participation with a collective’s activity should be understood as described above in Section 2. It is partly subjective: how far do their motives and goals, both explicit and implicit, overlap with the motives and goals driving other members of that collective to participate in the activity? It is also partly about objective causal contribution: how much of a difference did this individual make to the relevant result? And it is, thirdly, partly about causal interdependence: to what extent did this person making their contribution influence others towards making their own contributions, and vice versa?

For illustration, consider the statement ‘the people who got this candidate elected don’t value preserving such-and-such an institution’. Set aside for now the interesting question of what it is for a collective to value or not value something, and read this claim purely distributively, as

meaning that the people in question individually don't value preserving that institution. There might be a variety of attitudes towards the institution among the people who had some role in the election outcome. Some might actively want to weaken it; others might mildly dislike it; many voters probably haven't the foggiest idea what it is. On the other hand, at least some of the candidate's voters might have at least a vague attachment to the idea of it, and some political allies might consider it very important, but felt compelled to endorse the candidate anyway.

To determine the truth of the statement 'the people who got this candidate elected don't value preserving such-and-such an institution', take the set comprising all the former contributors, those who either don't care at all or are hostile to the institution. The statement will count as true if that set meets two conditions: (1) those people are a plausible candidate for being the collective agent who got that candidate elected, and (2) that set substantially overlaps all other plausible candidates.

Note that since 'substantially overlaps' can have higher or lower standards depending on context, the truth of the whole statement might be likewise contextual. Suppose the set of contributors who are indifferent or hostile to the institution is both large and contains many of the people most heavily involved in the candidate's election. Then it will likely fulfil condition 1, and fulfil condition 2 at least for moderate or low standards of 'substantially overlaps'. This would make the original statement true in casual conversation. But in a context with high standards for 'substantial overlap', such as an academic discussion of political sociology, it might be that this set no longer meets condition 2, so that the statement counts as false, a homogenising oversimplification. This seems to me an appropriate result.

5 Why Messiness Remains

Lewis thinks, and I agree, that manyism provides a basically successful resolution to puzzles of multiplication concerning humans and cats. Of course, the idea that each of us is really a cloud of trillions of almost-identical people is surprising and weird. But given the right semantic machinery, it need not be problematic for practical purposes, and arguably

this is the kind of weirdness we should expect from our dualistic intuitions crashing into the reality of us actually being divisible material objects.

By contrast, collective agents are just messier and less coherent entities than human individuals, and so manyism will not resolve the problem of overlapping collectives as fully and satisfactorily. Note in particular three key differences.

First, the plausible criteria for being part of a social collective vary much more widely than the criteria for being a part of a human individual. As well as various institutional criteria, there are all the varying degrees of participation described above, possibly specified in subtly different ways, whereas to be part of a human individual is usually defined in straightforwardly biological terms (though narrower psychological or neural terms might yield different ideas about whether parts of my body are really parts of me).

Second, for any given criterion, there are likely to be far more borderline members in a social collective than borderline parts of a human body. Skin cells and stray hairs might be neither determinately still part of a person's body nor determinately no longer parts, but the great bulk of the body is easy to identify as made up of definite parts. Whereas the collective involved in a large-scale action like oppressing a certain group might consist *primarily* of people with only middling degrees of participation, who might or might not be counted as participants depending on context.

Thirdly, the things which count as part of a human individual by one criterion but not by another are generally either insignificant (like stray hairs) or move around everywhere with the definite parts (like body parts, if the individual themselves is a mind or brain). Thus, the differences between different members of the cloud of human individuals can usually be ignored. By contrast, people who are members of one collective but not another may behave in importantly different ways, with importantly different motives and dispositions. The overlapping collectives that we might identify as the agent who got a given candidate elected differ in very significant ways. For instance, the set of ideologically committed voters and politicians is importantly different from the larger set of voters, in that one set is open to partially or wholly changing its collective mind while the other is, by definition, ideologically committed.

As a result of these differences, talk of ‘the agent’ of a diffuse collective action will be much messier than talk of ‘the agent’ of (what we would normally call) an individual human’s action. Fewer things will be determinately true about ‘the people who elected the president’ than are true of any individual person. This chapter has not sought to *remove* this messiness, but to clarify why it makes things difficult, and why that difficulty does not indicate any fundamental theoretical problem with viewing these phenomena as agential.

6 Conclusions

The conclusion I draw is a modest one: we probably cannot, in most circumstances, assign social responsibility in any much more precise or satisfying way than we currently do, but that does not mean that what we say is meaningless or confused. We should not shy away from saying that a large, dispersed group of people has actively done something important. But we should recognise such claims as shorthand for complex quantified claims: there are many large dispersed groups that did this thing, but they all substantially overlap with one another.

Notes

1. I will use the term ‘collective action’ and ‘collective agent’ deliberately loosely here, without commitment as to, for instance, whether a collective agent persists over time, or is simply some people who, on a certain occasion, do something collectively.⏎
2. Here I have in mind primarily *constitutive* social construction of objects and properties – see Hacking 1999; Haslanger 2003; Diaz-Leon 2013; Ásta 2015.⏎
3. See Effingham 2010, Ritchie 2013, for some discussion of the ontology of groups; I take no stand on what a group *is*, beyond noting that it has a set of members at any given time.⏎

4. For discussion see, e.g., Gilbert 1990, 2009; Bratman 1993, 2014; List and Pettit 2011; Neog 2020, 2023.⏎
5. See e.g. Collins 2019; cf. Neog 2023.⏎
6. https://www.reddit.com/r/Anarchy101/comments/1624lsp/what_is_the_origin_of_the_famous_utah_phillips/⏎
7. For influential early discussions see esp. Geach 1980; Unger 1980; Lewis 1993; Merricks 2001; for recent discussions see Sutton 2014, 2015; López de Sa 2014; Roelofs 2022.⏎

References

Ásta. 2015. ‘Social Construction’, *Philosophy Compass*, 10 (12): 884–892.⏎

Bratman, Michael. 1993. ‘Shared Intention’, *Ethics*, 104 (1): 97–113.⏎

Bratman, Michael. 2014. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.⏎

Collins, Susan. 2019. ‘Collective Responsibility Gaps’, *Journal of Business Ethics*, 154 (4): 943–954.⏎

Diaz-Leon, Esa. 2013. ‘What is Social Construction?’, *European Journal of Philosophy*, 23 (4): 1137–1152.⏎

Effingham, N. 2010. ‘The Metaphysics of Groups’, *Philosophical Studies*, 149: 251–267.⏎

Farrell, J. 2017. ‘When a Candidate Conspired With a Foreign Power to Win An Election’, *Politico*. https://www.politico.com/magazine/story/2017/08/06/nixon-vietnam-candidate-conspired-with-foreign-power-win-election-215461/ (Accessed July 2025).⏎

Geach, P. T. 1980. *Reference and Generality*. 3rd ed. Ithaca, NY: Cornell University Press.⏎

Gilbert, M. 1990. 'Walking Together: A Paradigmatic Social Phenomenon', *Midwest Studies in Philosophy*, 15: 1–14.⏎

Gilbert, M. 2009. 'Shared Intention and Personal Intentions', *Philosophical Studies*, 144 (1): 167–187.⏎

Hacking, I. 1999. *The Social Construction of What?* Cambridge: Harvard University Press.⏎

Haslanger, S. 2003. 'Social Construction: The "Debunking" Project', in F. Schmitt (ed.), *Socializing Metaphysics*, pp. 301–325. Lanham: Rowman & Littlefield.⏎

Lewis, David. 1993. 'Many, but Almost One', in John Bacon, Keith Campbell, and Lloyd Reinhardt (eds), *Ontology, Causality, and Mind: Essays in Honour of D. M. Armstrong*, pp. 23–45. Cambridge: Cambridge University Press.⏎

List, C. and P. Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.⏎

López de Sa, Dan. 2014. 'Lewis vs Lewis on the Problem of the Many', *Synthese*, 191 (6): 1105–1117.⏎

Martens, Judith, and Roelofs, Luke. 2018. 'Implicit Coordination: Acting Quasi-Jointly on Shared Implicit Intentions', *Journal of Social Ontology*, 4 (2): 93–120.⏎

Merricks, Trenton. 2001. *Objects and Persons*, New York: Oxford University Press.⏎

Neog, Bhaskarjit. 2020. 'Metaphysics of Group Moral Responsibility', *Journal of Human Values*, 26(3): 238–247.⏎

Neog, Bhaskarjit. 2023. *What Responsibility? Whose Responsibility?: Intention, Agency, and Emotions of Collective Entities*. London: Routledge India.⏎

Ritchie, Katherine. 2013. ‘What are Groups?’, *Philosophical Studies*, 166: 257–272.⏎

Roelofs, Luke. 2022. ‘No Such Thing as Too Many Minds’, *Australasian Journal of Philosophy*, 102 (4): 1–16.⏎

Simon, Jonathan. 2017. ‘The Hard Problem of the Many’, *Philosophical Perspectives*, 31 (1): 449–468.⏎

Sutton, C.S. 2014. ‘The Supervenience Solution to the Too-Many-Thinkers Problem’, *Philosophical Quarterly*, 64 (257): 619–639.⏎

Sutton, C.S. 2015. ‘Almost One, Overlap and Function’, *Analysis*, 75 (1): 45–52.⏎

Unger, Peter. 1980. ‘The Problem of the Many’, *Midwest Studies in Philosophy*, 5: 411–467.⏎

Zimmerman, Dean W. 2010. ‘From Property Dualism to Substance Dualism’, *Aristotelian Society Supplementary* Volume, 84 (1): 119–150.⏎

OceanofPDF.com

9 Team Reasoning Without Assurance

Morality as a (Largely) Non-Coordinating Practice

Jeremy Randel Koons

DOI: 10.4324/9781003612827-10

1 Introduction

The literature on game theory recognizes different kinds of coordination games. Choosing which side of the road to drive on is often given as an example of a *pure coordination game*: There are two pure Nash equilibria, and it doesn't matter which side the two players choose, as long as they choose the same side. However, there are also *assurance games* such as the stag hunt. In an assurance game, the higher payoff can only be achieved if both players cooperate. Thus, each player should cooperate only if the other does (or seems likely to); but otherwise, defection is the rational choice.

Game theoretic approaches to moral reasoning have advanced significantly in recent years with the proposal of team reasoning as a valid model of reasoning. However, these approaches almost universally start with consequentialist presuppositions; such reasoning inexorably leads to the conclusion that morality (if justified game-theoretically) must be a species of assurance game.

Unfortunately, if morality is an assurance game, then it will be difficult to sustain a deontological reading of moral obligation. For deontology will often require us to perform the morally obligatory action even when we

know others will not; it is difficult to make sense of such an obligation within a normative practice which is itself a type of coordination game.

I will argue that understanding team reasoning in consequentialist terms is not mandatory. Drawing on the work of Wilfrid Sellars, I will characterize a deontological form of team reasoning—Principled Team Reasoning. While all team reasoning has a coordinating element to it, I will argue that our moral practice largely does *not* serve a coordinating function, and cannot be understood as an assurance game. It can be rational—and even obligatory—to act non-cooperatively per the dictates of morality.

2 (Almost) Purely Coordinating Practices

A purely coordinating practice is one justified purely because we need to coordinate our behavior: It is the coordination itself that serves some goal. The idea of a purely coordinating practice is an ideal, but there are a number of close approximations. Rules about which side of the road to drive on, for example, serve a purely coordinating purpose. It doesn't matter if people drive on the left or the right side; it just matters that people agree on a convention—it is the *coordination* that matters. Plausibly, such practices can be modeled as *pure coordination games*.

Purely coordinating practices have features making them ill-suited for certain normative purposes. Because their main purpose is to coordinate behavior, the rules of such practices are *reified*, meaning they are often institutionalized, and they are also tied to a particular place and time and can display a high degree of relativity (both historical and intercultural). For example, Sweden switched from driving on the left-hand side of the road to driving on the right-hand side of the road literally overnight, on September 3, 1967. As is the nature with purely coordinating practices, it was correct to drive on the left-hand side of the road on September 2, and incorrect on September 3. We do not criticize left-side drivers from the 1950s and argue that they were doing something objectively 'wrong'; and once the rule was changed, it changed instantly and for everyone. This overnight change is possible because such rules are institutionalized—the change was proposed by the Swedish Prime Minister, approved by the Swedish Riksdag, and

overseen by the Swedish Commission for the Introduction of Right-Hand Driving.

Additionally—and this is a familiar point—the rules of a purely coordinating practice are to a large degree arbitrary. It doesn't matter which side of the road people drive on, as long as there is agreement. Moral rules, by contrast, are not reified; not institutionalized; and not arbitrary in the ways described above. Thus, while purely coordinating practices have their use, we should be reluctant to characterize morality as such a practice.

3 Partially Coordinating Practices

Other of our normative practices resist game-theoretic treatment because they do not serve a wholly coordinating function. Some such practices are Janus-faced, in that they serve two goals which may often be in tension. A prime example is our epistemic practice—our concrete practice of assigning entitlement or justification to beliefs, assigning more or less epistemic authority to various agents or institutions, designating various methodologies are justification conferring and others as irrational or unreliable. Obviously, our epistemic practice serves various epistemic goals —truth, or explanation, or what have you. (I am not interested in taking a stand on what precise epistemic good(s) structure(s) our epistemic practice.) Less obviously, our epistemic practice also serves an important social goal, namely, *commitment coordination*. I have argued for this conclusion at length elsewhere (Koons 2025); I will only briefly recapitulate some of my argument's main points here.

3.1 Epistemology as a Partially Coordinating Practice

It is uncontroversial that the construction of human society and its institutions requires large-scale commitment coordination. As Berger and Luckmann write, 'I live in the common-sense world of everyday life equipped with specific bodies of knowledge. What is more, *I know that others share at least part of this knowledge, and they know that I know this*. My interaction with others in everyday life is, therefore, constantly affected by our common participation in the available social stock of knowledge' (Berger and Luckmann 1966: 56, emphasis added). Tomasello echoes this

thought, writing, 'For survival, modern human groups had to begin operating as relatively cohesive collaborative units, with various division-of-labor roles (see Wilson 2012). This created the problem of how individuals could coordinate with in-group strangers, with whom they had no personal common ground. The solution was the conventionalization of cultural practices: everyone conformed to what everyone else was doing, and expected others to conform as well (and expected them to expect them to, etc.), which created a kind of cultural common ground that could be assumed of all members of the group' (Tomasello 2014: 138). Indeed, the very ability to cooperate and exercise collective intentionality—in the form of joint intentions, shared beliefs, collective acceptance, or other shared intentional states—requires an extended ontogeny during which children are inducted into a complex system of norms and institutions developed over centuries and generations: '[S]kills of collective intentionality are not innate or maturational…[T]hey are biological adaptations that come into existence only through an extended ontogeny in a collectively created and transmitted cultural environment' (Tomasello 2014: 147). Searle (2009), Tuomela (2002), and others have also written extensively about how basic human institutions (such as property, money, etc.) are created and sustained by shared commitments or other attitudes, which presuppose some kind of coordination of commitments around the elements of these institutions. Thus, an indefinitely large range of shared statuses (knowings-that and knowings-how)—as well as institutions and other social structures—underlie the very possibility of human community.

However, the same is true not merely of social or institutional commitments, but also of beliefs about ordinary empirical facts—the sort of facts we are concerned with when we talk about the locations and alethic possibilities of objects in space and time, and the sort of facts investigated by the natural sciences. The scope of commitment coordination required for any large-scale cooperation project—such as the Large Hadron Collider, or engineers collaborating on the design of a new car, or what have you—is immeasurable. Granted, not every collaborator will share every commitment—there is an epistemic division of labor, which is part of why you have collaborators in the first place—but an indefinite range of shared commitments, including commitments about whose entitlements to defer to

on issues one doesn't know anything about, necessarily structure such collaboration.

To consider a concrete example of large-scale social coordination enabled by commitment coordination, consider international efforts to eradicate diseases like smallpox and polio. The World Health Organization (WHO) began its Intensified Eradication Program for smallpox in 1967, by which point the disease had already been eliminated from North America and Europe. In short order, smallpox was eliminated from South America, Asia, and finally from Africa in 1977 (US Centers for Disease Control 2024). The success of this eradication effort required smaller, more specialized pools of common belief and knowledge—such as how to create vaccine supply chains, how to implement 'ring vaccination' when an outbreak is detected, and so on—and also the corresponding practical commitments. But such success also requires more *global* commitment coordination. It requires that the population generally be made knowledgeable about the dangers of the disease; the existence, safety, and effectiveness of the vaccine; measures that can be taken to avoid infection in the first place; etc. Further, the general population needed to coordinate regarding the relevant practical commitments—they needed to form (and execute) the intention to get the vaccination, to get their children vaccinated, to follow whatever prevention and quarantine measures were prescribed by the health authorities, and so on.[1] The success of the eradication effort depended upon the success of this commitment coordination. That this is true can easily be seen by the failures of the current efforts against COVID-19—driven largely, one might argue, by disinformation campaigns that inhibit the commitment coordination necessary for a successful fight against the pandemic.

I conclude that commitment coordination is essential for the existence of human community, and for the success of any kind of cooperation, and indeed for much of our rational practical activity. Once we understand this, we can also see that a central social function of any society's epistemic practice must be *commitment coordination*—that is, the production of consensus sufficient to enable community, communication, and fruitful coordination.

3.2 Reification Within Partially Coordinating Practices

Partially coordinating practices partially reify, and display some degree of relativity, making them an inappropriate model for our moral practice. Our practice of epistemic evaluation, for example, is relative not only to historical era but also (perhaps) to other features of a person's epistemic situation.

Thus, we might say of an astronomer living in the 2nd century CE that it was rational for them to believe in the geocentric nature of the universe: Their belief was of course false, but no specifically *epistemic* criticism could be raised against their doxastic conduct. Thus, we do relativize epistemic evaluation to an extent.

But we do not fully relativize it. Thus, we do not simply defer to different epistemic communities (especially contemporary communities) and deem as correct whatever epistemic evaluations they happen to endorse. Thus, we might encounter a Facebook group dedicated to vaccine refusal or flat eartherism, and deem their beliefs irrational and unjustified. It matters not that they are all in agreement, and form a coherent community—they are still subject to specifically epistemic criticism. Thus, epistemic evaluation does not relativize to the extent that traffic rules or linguistic rules do. We hold communities somewhat accountable to what we take to be objective standards.[2]

Part of what explains this feature of our epistemic practice is, again, its dual function. Our epistemic practice is structured by two distinct goals which may be in tension. First, our epistemic practice serves purely epistemic goals, such as truth or explanation or what have you. This pushes our epistemic evaluation in a non-relativistic direction. But our epistemic practice serves a distinct coordinating function,[3] and this is part of what pushes us to relativize the practice along certain dimensions. Given the way in which partially coordinating practices relativize, we should be reluctant to view morality as such a practice.

4 Largely Non-Coordinating Practices

Most philosophers are reluctant to relativize morality to different communities, or even to different historical eras. Thus, slavery is wrong

even in a community where it is widely-accepted. To the extent that we do relativize, we relativise *epistemically*, not *morally*. Thus, we say that while a historical practice was objectively morally wrong, its participants might be excused because their moral ignorance was understandable. We might put this by saying that morality displays *epistemic* relativity but not *content* relativity.

Thus, moral rules do not relativize; and they do not reify. Driving on the left side of the road might be permissible yesterday but not today. But there is no mechanism whereby slavery will overnight go from moral to immoral or vice versa. Part of this is because of the non-institutional character of morality. But part of this is because coordination is not morality's chief purpose. If it were, then it would be important for people to converge on a rule (R1), perhaps even if this rule were non-optimal. It would also be ideal if later this rule were replaced with an optimal one (R2). Thus, morality would relativize in the same way that epistemology does—a conclusion we should resist.

4.1 Coordinating Practices as Subparts of Morality

One might argue that while morality is itself a largely non-coordinating practice, it involves subpractices that are themselves partially or largely coordinating. This may be true. But it is easy to overstate; it also does not demonstrate that morality is *in general* a coordinating practice.

The most obvious examples of practices that are (a) coordinating and (b) seem to impose specifically moral (as opposed to customary) obligations involve situations where a certain threshold of participation is required to achieve a social benefit. One such example is that of vaccination. Various individuals in a community are not or cannot be vaccinated—newborns, people who are immune compromised, etc. Thus, to prevent a vaccine-preventable disease from circulating in a community, a certain threshold of vaccinations must be reached—so-called 'herd immunity'. For highly-contagious diseases such as measles, this threshold is 95% of the population; it is lower for other viruses. Plausibly, those who can safely be vaccinated are obliged to do so in order to help the community reach herd immunity and achieve the corresponding communal benefit.

However, I am skeptical that other alleged examples of coordinating moral subpractices hold up to critical scrutiny. A category of situations often thought to create moral obligations involve ‘tragedies of the commons’. Consider, for example, our alleged moral obligation to act individually to reduce our carbon footprint. Supposedly, if each of us were to act independently, we could avert or significantly mitigate climate change. Or consider another oft-cited example of the tragedy of the commons: An overtaxed fishing ground, and the supposed obligation this creates for each angler to limit their catch in order to preserve this common resource.

While there may well be ‘tragedies of the commons’ that generate coordinating agreements which themselves generate or constitute moral obligations, I suspect these are rarer than is generally thought. For example, it is not clear what kind of individual obligation we have to reduce greenhouse emissions. It has been noted, for example, that statistically speaking, individuals are responsible for a small fraction of global greenhouse gas emissions; large corporations and state entities are by far the largest emitters: Since 1988, 100 corporate and state entities have been responsible for 71% of global greenhouse gas emissions, and 25 producers have been responsible for 51% of such emissions (Griffin 2017: 8). Trying to make emission mitigation a matter of individual responsibility—or even the collective responsibility of private citizens—illicitly shifts the burden away from those who are in fact causally and morally responsible for the problem.

Other cooperative systems for commons-management look less like systems of moral obligation and more like quasi-bureaucratic resource management institutions. Elinor Ostrom, the Nobel Prize-winning economist who has done as much as anyone to explain how cooperative institutions work, has detailed how groups come together to protect common resources—and what features these cooperative arrangements need to have in order to be self-sustaining. The most successful ones are institutionalized in various ways and have set procedures for monitoring compliance and sanctioning violation. Consider, for example, the system put in place in the 1970s by fishers in Antalya, Turkey, to preserve their fishery, whose viability was threatened by (among other things) overfishing. There is no need to describe the intricate system in detail; it suffices to

outline some of its key features. A list of the local fishers is prepared at the beginning of the season. All fishing locations are also listed. In September, there is a lottery that randomly assigns fishers to fishing locations. Thus, you might be assigned a good location or a bad one; it is a matter of chance and luck. But you are not stuck with this location for the whole season:

> From September to January, each day each fisher moves east to the next location. After January, the fishers move west… The system has the effect of spacing the fishers far enough apart on the fishing grounds that the production capabilities at each site are optimized. All fishing boats also have equal chances to fish at the best spots. Resources are not wasted searching for or fighting over a site…
>
> The list of fishing locations is endorsed by each fisher and deposited with the mayor and local gendarme once a year at the time of the lottery…Cheating on the system will be observed by the very fishers who have rights to be in the best spots and will be willing to defend their rights using physical means if necessary. Their rights will be supported by everyone else in the system. The others will want to ensure that their own rights will not be usurped on the days when they are assigned good sites. The few infractions that have occurred have been handled easily by the fishers at the local coffeehouse.
>
> (Ostrom 1990: 19–20)

The agreements created by this system might *create* moral obligations (just as in general, voluntary agreements to do morally neutral or morally good actions create *prima facie* obligations), but they are not *constitutive* of moral obligations. Systems like the one instituted by the fishers have many features that make them differ from systems of moral norms: They are formal and institutionalized and do not capture the non-institutional

character of morality; they are very local and highly convention-dependent; they are strongly subject to reification; etc.

I conclude, then, that we should not understand morality as a coordinating practice. There might be some very limited moral sub-practices that serve a coordinating function, but even this is easily (and often) overstated. This conclusion will have a variety of consequences for our attempt to understand morality as a *principled* form of team reasoning.

4.2 Morality and Self-Interest

One consequence of understanding morality as a largely non-coordinating practice is that we cannot justify our moral practice in terms of self-interest. To see why this is so, let us begin with our assumption that moral reasoning must be principled: We must be able to make sense of the fact that morality may require us to act non-cooperatively. The most promising attempts to justify morality in terms of self-interest appeal to mutualism, 'in which we all benefit from our cooperation but only if we work together' (Tomasello 2009: 52).[4] Notice, however, the 'only if'—if others do not cooperate, then it is not in my self-interest to cooperate. And so, if self-interest is what justifies my cooperation, then I should not cooperate, if others are also not willing. Thus, it is difficult to see how an appeal to self-interest can justify principled action.

To be sure, various authors have tried to find ways around this—for example, by suggesting that I signal I am a reliable partner by exhibiting cooperative behavior even in the face of non-cooperation—but (aside from my overall pessimism about the prospects of justifying morality by appeal to self-interest) I will set these attempts aside, as I think there is a potentially deeper worry about this approach. On this approach, there aren't actually any moral reasons *per se*—morality is wholly reducible to self-interest; and in cases where one's self-interest is not best-served by following the applicable moral norm, then one has no reason to do so. There are no distinctively *moral* reasons which might weigh in the balance against prudential or other sorts of reasons.

This is consistent with a long philosophical history of skepticism about moral reasons. This skepticism is puzzling: philosophers are in general not skeptical of other sorts of non-prudential reasons. For example,

philosophers are not skeptical of (say) epistemological or semantic reasons. Nor is it plausible to think that these reasons can be reduced to prudential reasons: To be able to consider the sort of prudential reasons that would justify the adoption of epistemic or semantic reasons, you would already have to abide by epistemic or semantic reasons! So prudential reasons cannot be prior—in the justificatory order—to these reasons; they cannot reduce to self-interest.

It is likely that philosophers hold that these other sorts of reasons are transcendentally necessary in a way that moral reasons are not. One cannot help but abide by (say) semantic reasons; whereas one need not abide by moral norms. They are optional in a way that semantic norms aren't. But this isn't true—one need not abide by semantic reasons. Animals don't, and get on just fine. What is true is that you must abide by semantic reasons *in order to be a certain sort of agent*. Thus, semantic reasons have a certain kind of *categorical validity*—they are rationally non-optional in a very specific way. The same may well be true of moral norms—not that one cannot act or be an agent without being bound by them, but that one is cut off from a specific valuable type of agency.

Thus, I am inclined to argue that an adequate moral theory has to claim that there are actually moral reasons. Theories that reduce moral norms to mutualistic cooperation do not create or acknowledge such reasons. Thus, we will not find our justification for morality here.

5 Constitutive Agreement and Moral Reasoning

We are, it seems, bound by two incompatible requirements. On the one hand, we want to make sense of morality as a species of team reasoning, and team reasoning must involve agreement or mutual commitment or something similar. On the other hand, we have insisted that an agent might legitimately hold—and act on—an unshared moral commitment. If we want to resolve this apparent contradiction—if we are to understand moral norms as (a) having a deontological character while also (b) resulting from a species of team reasoning—then we must understand constitutive agreement and the role it plays in our normative practices.

What is constitutive agreement? In ‘On Reasoning about Values’ (ORAV), Sellars gives the following example using the fictional ‘Whooping Crane Society’ (WCS):

> Shall [each of us WCS members promote the survival of whooping cranes]
>
> is reasonable *for members of the WCS*, because it is by virtue of that shareable intention that there is such a thing as the WCS.
>
> This intention defines what might be called “the WCS member’s point of view.” It is the prime mover intention of this point of view, and generates, given matters of factual implications and premises, subordinate intentions, the relative reasonableness of which is grounded in its intrinsic and non-relative reasonableness as constituting the WCS point of view.
>
> (ORAV: §§186–7, footnote deleted)

Sellars writes that the WCS exists ‘by virtue of that shareable intention’, but of course the implication is much stronger. The intention in question must not merely be shareable; the WCS can only exist if the intention is *in fact* shared. This is what is meant by constitutive agreement: A particular normative practice exists—is constituted by—a set of shared intentions or other attitudes. Whatever ‘subordinate intentions’ one might have *qua* member of the WCS— ‘Shall_{we} [I petition City Council to preserve the wetlands within the city limits]’—are only reasonable (and I can only intend them *qua* WCS member) because of this original constitutive agreement. (Sellars has a detailed view about the relationship between constitutive agreement and these subsidiary intentions, a view we will discuss in Section 8.)

A further feature of constitutive agreement—derivative from the fact that it constitutes communities—is that it makes certain kinds of agreement and

disagreement possible. You and I might disagree about the treatment of whooping cranes, but we can only disagree (or agree) *qua* members of the WCS to the extent that we both endorse the founding intention of the Whooping Crane Society (and are thereby members of the WCS).

The source and nature of a practice's constitutive agreement will depend on the type of practice under discussion. Constitutive agreement in almost purely coordinating practices may be institutional in origin, and will (as discussed earlier) be highly reified: the agreements constituting rules about what side of the road to drive on are generated by a central authority and can be changed quite literally overnight.

One additional note about constitutive agreement: The word 'agreement' (along with the WCS example) might mislead the reader to think that constitutive agreements are voluntary. But 'agreement' as used here has more the sense of 'accord' or 'conformity'. Thus, for example, communication among speakers of Thai requires that these speakers largely agree (= exhibit conformity) regarding the meanings of words in Thai, Thai grammar, etc. But it would be absurd to say that people voluntarily agree to enter a particular linguistic scheme, as one is not a rational agent prior to being socialized into a language. Thus, constitutive agreement need not be the sort of agreement that could even in principle be voluntary.

Constitutive agreement always serves the goal of coordination to some extent. In some practices (e.g., linguistic norms or rules of the road), constitutive agreement is almost wholly directed at coordination. In other cases of constitutive agreement, the connection is looser: With the WCS, the constitutive agreement creates the society and directs its members towards a common goal—so there is an obvious element of coordination—but members might disagree over the means of achieving that goal while nevertheless intending various means *qua* members of the WCS. This is why I describe morality as a *largely* non-coordinating practice. No social practice gets off the ground without an initial coordination of attitudes. But given this background of attitude-coordination, one can still distinguish between practices that serve largely to coordinate (such as norms surrounding which side of the road to drive on) and practices that largely do not (such as morality).

Sellars suggests that the moral community is created by a constitutive agreement analogous to the one that creates the WCS:

> Is there an intrinsically (and hence categorically) reasonable formal intention which is the regulative principle of the moral point of view?
>
> Analogy would suggest the following:
>
> Shall [each of *us* do that which, in the circumstances, promotes the happiness of each and every one of *us*, all relevant things considered].
>
> (ORAV: §§193–4)

How do we get from this very general moral exhortation to concrete moral maxims? For Sellars, 'ethical truths are the projections of "matter-of-factual" truths…into the framework of intentions and purposes' [Science and Metaphysics (SM): Ch. 5, §3/117]. Thus, we take a variety of nomologicals connecting the promotion of the happiness of each and every one of us to the doing of particular actions in particular circumstances, and then we reason as follows:

> Shall be [each and every one of us leads a satisfying life, all relevant things considered]
>
> 'Each and every one of us leads a satisfying life, all relevant things considered' implies 'Each of us does A_1 in C_1, A_2 in C_2, etc'.
>
> Therefore, Shall [Each of us does A_1 in C_1, A_2 in C_2, etc.]
>
> Therefore, Shall_{we} [I do A_1 in C_1, A_2 in C_2, etc.]

Presumably, due to various disagreements (concerning the contents of a satisfying life, concerning the correct nomologicals to get us from the intrinsically valid intention to concrete actions, etc.), individual people

might disagree about what their obligations are in particular situation: Smith might argue that one ought to do A_1 in C_1, whereas Jones argues that one ought to do A_2 in C_1. But given the derivation, both of these 'ought' statements will be expressed in the same way:

$$\text{Shall}_{we}\ [\text{I do } A_i \text{ in } C_i]$$

That is, Smith and Jones are intending from the group point of view, even if they disagree. (In fact, they only disagree *because* they are both intending from the group point of view—Sellars insists on this.)

Thus, Sellars's model of moral reasoning gives us a hint at how morality could be a species of team reasoning while allowing for non-shared attitudes and actions. However, what Sellars offers here is only a sketch. To make the case that morality is a case of team reasoning—specifically, a kind of deontological team reasoning—we must first give a more precise characterization of team reasoning. This will, in turn, help sharpen our characterization of constitutive agreement—and help us see the sense in which constitutive agreement does and does not necessarily involve coordination.

6 Team Reasoning

What is team reasoning, in its most general form? To engage in team reasoning is to reason in the we-mode (in Tuomela's strong sense), where one's reasoning makes essential use of group attitudes (such as group intentions or group beliefs/commitments). Thus, consider the following attitude which Smith can have qua member of Organization:

Smith: Jones is President of Organization.

Smith's belief is properly characterizable as at attitude_{Org} given two conditions: First, Smith and other members of Organization are mutually committed to various rules about the structure of Organization (e.g., that there is a leader, about the relation of the leader to ordinary members, etc.), about the bylaws (voting procedures, procedures for conducting meetings,

rules of quorum, rules for moving and seconding motions, etc.), and so on; and second, the belief in question is inferentially derived from some set of these mutual commitments (in a manner to be outlined in Section 7). I say ‘commitment’ because this is more accurate than ‘belief’: I may not recall off the top of my head what the rules of quorum are, but I know that the rule is laid out in the bylaws, and am committed—to a greater or lesser extent—to binding myself by these bylaws (as are other members of Organization).

Given the notion of a group attitude, we can also readily understand the general notion of team reasoning. Such reasoning can involve doxastic attitudes such as beliefs as well as practical attitudes such as commitments:

> We at Organization believe that climate change is a threat to our community’s future.
>
> Organization shall reduce its carbon emissions by 50% over the next ten years.

For reasons described in the previous paragraph, both of these attitudes are characterizable as attitudes$_{Org}$, and so reasoning involving them involves reasoning from a group point of view. We can also of course engage in group practical reasoning:

> Shall$_{we}$ end the corrupt and incompetent reign of Jones.
>
> Cooper is honest and capable (and available).
>
> Therefore, Shall$_{we}$ elect Cooper as the new president of Organization.

7 Constitutive Agreement and Unshared Team Attitudes

The concept of constitutive agreement allows us to make sense of the notion of unshared group attitudes—something of crucial importance if we

are to articulate a conception of deontological team reasoning. Let us begin by noting that as Tuomela points out, even in the case of 'strongly social concepts', like that of being a leader (i.e., in an organization), there are 'constraints for correct collective acceptance' (2002: 147). Thus, if Smith is elected leader of Organization, but it turns out that a quorum wasn't present at the meeting, then the election was not valid, and he is not in fact the valid leader, *even if people take him to be so*. Therefore, someone could challenge his status as leader, demand a re-vote, etc. Thus, I may be the only person in Organization to hold the attitude

B: I Believe$_{Org}$ [Smith is not the valid leader]

and yet my group-attitude is correct—I am *entitled* to this attitude—though this attitude is not shared by the organization as a whole.

There are a few crucial things to note about this case. The position 'elected leader of Organization' only exists because of various *constitutive agreements* that obtain among members of Organization. Thus it depends on agreement about the structure of Organization (e.g., that there is a leader, about the relation of the leader to ordinary members, etc.), but also on a large body of what I will call *de re* agreements: We agree (explicitly or implicitly) that the bylaws of Organization are binding when it comes to elections, and thereby (*ceteris paribus*) agree *of* the contents of the bylaws (voting procedures, procedures for conducting meetings, rules of quorum, rules for moving and seconding motions, etc.) *that* they will govern elections.

Next (and perhaps most crucially for our purposes) I am only able to entertain belief B *qua* member of Organization, and am only able to disagree with other members of Organization *qua* members of Organization, because of the constitutive agreements that enable such beliefs and such disagreements. Thus, I can hold belief B (and other members of Organization can disagree) only because we agree about the structure of Organization (e.g., that there is a leader, about the relation of the leader to ordinary members, etc.), about the bylaws (voting procedures, procedures for conducting meetings, rules of quorum, rules for moving and

seconding motions, etc.), and so on. Constitutive agreement underlies the very possibility of agreement and disagreement.

This is not merely true in institutional cases; this is true wherever we dispute over norms. A dispute over a semantic norm (e.g., the proper meaning of 'beg the question') presupposes massive constitutive agreement —concerning the meanings of the constituent words 'beg' and 'question', how words combine meaningfully, the meanings of the sentences we are using to debate the question, etc.—without which we couldn't even be characterized as having a *semantic dispute*. Nor could we have a specifically moral argument without a background of constitutive agreement about the subject matter of morality. Any discourse—including dissent—presupposes the kind of shared group attitudes that are at the heart of Tuomela and Gilbert's accounts. And when Sellars argues that Jones can have a moral we-intention that no one else possesses, one should understand this in the context of an ongoing social practice with significant constitutive agreement in the background. So, suppose that Jones has a moral attitude ($Shall_{we}$ [I do A_i in C_i]) that others do not share. Essentially, Jones is making (in the moral sphere) the same move I made in Organization: Jones is conveying that given this background agreement (say, agreement on the formal end of morality, certain abstract principles of justice or fairness, and so on), we are *in fact* committed to doing A_i in C_i. Such a move can only be understood in the context of significant agreement over what morality requires, what can be appealed to as reasons, and so on. In general, the game of giving and asking for reasons presupposes general agreement in attitudes, as a background for arguing over the remaining areas of disagreement.[5] Thus, while PTR allows one to have a non-shared moral attitude, such an attitude does presuppose a significant constitutive agreement in the background. Thus, the team is never wholly absent from team reasoning in principled team reasoning. Sellars's view (laid out in ORAV) is that what makes *my* attitude a we-attitude is not that it is in fact shared, but that it is *inferentially downstream* from some appropriate constitutive agreement or set of constitutive agreements. Route of derivation (not actual sharedness) is what makes for a we-attitude—but such attitudes essentially depend on constitutive agreement.

Elsewhere (Koons 2022), I offer a more formal analysis (based on Tuomela's notion of an *ethos*) of what it is to have a group attitude that allows for unshared group attitudes. Tuomela defines an ethos as follows: 'The *ethos* of a group g in its strict sense is defined as the set of *constitutive* goals, values, beliefs, standards, norms, practices, and/or traditions that give the group motivating reasons for action' (Tuomela 2007: 16). So, the ethos of Organization will be—at least in part—the rules constitutive of Organization, such as the charter, bylaws, and other rules that make organization exist, and in virtue of which we can be said to possess anything characterizable as $\text{Attitude}_{\text{Org}}$. Group cohesion—indeed, group constitution—requires 'the collective acceptance of E [the ethos] (involving collective commitment)' (Tuomela 2007: 32). We might say that I am committed to an ethos *qua* member of a group iff:

(a) I am committed to the ethos;
(b) I have an expectation that other members of the group are committed to the ethos; and
(c) I have an expectation that other members of the group have an equivalent expectation; i.e., an expectation that I am committed to the ethos.

By 'I am committed to the ethos' I mean simply 'I am committed to the rules, procedures, practices, etc. constitutive of the ethos'. I am eliding a number of no doubt thorny difficulties, as practically speaking, no two agents ever have *exactly* the same set of commitments regarding any set of norms, be they semantic, moral, epistemic, or whatever. But these incommensurability worries are not unique to this type of account; and in any case, what I am trying to offer here is an account of how an attitude can still count as a group attitude even if it is *not* shared—simply on the basis of existence of constitutive agreement. So, this account relies merely on the fact of rational agents sharing a 'form of life', even if precisely what is shared will vary from individual to individual.

Given that agents share an ethos, what is it for a non-shared attitude based on this ethos to be a we-attitude in the strong sense? For example, a

non-shared commitment that-P (say, that Smith is not the valid leader of Organization) will be a strong we-mode commitment iff, in addition to (a)-(c) holding:

> (d) I have a commitment to P, which commitment is epistemically based on the ethos (plus perhaps on certain other factual claims to which I am entitled).

The resulting notion of an attitude qualifies as a we-attitude in the strong sense.[6] It is certainly not a strictly *individual* commitment; it is one I hold *qua* member of Organization. But it is stronger than what Tuomela would call a 'weak we-mode' attitude. In the weaker case, Tuomela writes that acting as a group member in the weak sense 'is based on the group members' sharing the ethos but being only privately (and not collectively) committed to it' (2007: 29). Agents in the weak case do not act *because* of a joint commitment. They act on *private* commitments. But my belief B is not held for private reasons—it is held for reasons essentially tied to the ethos of Organization (and indeed for reasons that cannot even be *articulated* without reference to the ethos, or constitutive rules, of Organization). It is a belief I hold *qua* member of Organization. It is essentially a group belief. It is much more plausibly characterized as an attitude in the we-mode in the strong sense, rather than in the weak sense.

Next—and perhaps most importantly—to the extent that commitment P is epistemically based on the ethos (plus perhaps on certain other factual claims to which I am entitled), commitment P represents an *entitlement* that I have achieved. And just like $\text{Belief}_{\text{Org}}$, such entitlement is a *we-mode entitlement*, as its epistemic status derives from the ethos (plus other facts whose relevance and status is determined by the ethos—*no* facts have an interpretation *except* in light of an ethos); and my entitlement to P places corresponding obligations on other members of the community, as discussed in Section 5 above. Thus, I can conclude with the following definition: I will have strong we-mode entitlement to a non-shared commitment that-P (say, that Smith is not the valid leader of Organization) iff, in addition to (a)-(d) holding:

(e) The ethos (plus perhaps on certain other factual claims to which I am entitled) *entitle* me to commitment P.

But because we, *qua* members of a group, are mutually committed to the ethos; and because my commitment P counts as a we-mode commitment; and—most importantly—because per clause (e) I possess entitlement$_{group}$ to P, this entitlement extends to other members of the group. It is binding on them, whether they recognize it or not.

The upshot of this discussion is that this notion of constitutive agreement allows for deontological style *reasoning*, in that I can reason about what we are all committed to even when others disagree. If we think of the intention constituting the moral community as that community's ethos, then I can reason from that ethos to conclusions that are in fact binding on the group, whether or not the group recognizes this. 'Shall$_{we}$ (I do A_i in C_i)' might in fact follow from the constitutive intention of the moral community; and if it does, then it is an intention all ought rationally to endorse. I may be correct so to intend—even if others do not.

8 Unshared Attitudes vs. Unshared Behavior

At this point, it is perhaps natural to point out that while it is one thing to have an unshared attitude *qua* member of Organization, it is another thing to engage in unshared behavior *qua* member of Organization. I may not *believe* that Smith is the leader of Organization. But shall I then also *act* as though he is not leader of Organization? That is clearly a separate question; and the answer is not as clear-cut. Nevertheless, given the account offered above of unshared group attitudes, unshared group intentions seem conceptually coherent (if not always prudent). Thus, one is free to act as though Smith is not leader of Organization (at the possible risk of getting fired).

The conceptual coherence of unshared group intentions is (part of) what we need for our deontological conception of morality. But we also need to show that such unshared group intentions are not merely coherent, but sometimes *rational*—or even *obligatory*. Let us see how Sellars tries to make sense of this.

8.1 Sellarsian Teleology

For Sellars (as for Kant), practical reasoning is means-end reasoning (although Sellars's view of practical reasoning diverges from Kant's at a number of crucial points). Sellars argues that while an intention cannot be true, it can be reasonable. For Sellars, if an end-intention (e.g., saving the whooping cranes) is categorically reasonable, then the means-intention (petitioning the city council to preserve the wetlands) is also reasonable. Crucial to Sellars's metaethical project is this: If there were an end-intention that were *intrinsically* reasonable, then it would confer categorical (albeit derivative) reasonableness on all subsidiary intentions which can be derived from it. This is part of Sellars's strategy to, as he puts it, "[harmonize] 'teleological' and 'deontological' themes…in Kant's ethics" [Obligation, Intersubjectivity and the Moral Point of View (OIM): §147].

Now, this talk of 'teleological themes' might encourage a consequentialist reading of Sellars here, and put us back to square one. And to be sure, for most of his career Sellars is probably best read as a rule utilitarian. In OIM, Sellars suggests that the intrinsically valid intention structuring the moral point of view is the following:

> It shall$_{we}$ be the case that our welfare is maximized
>
> (OIM: §120)

The principle explicitly mentions maximization, and his discussion elsewhere in the same piece (see especially OIM: §85) suggests consequentialist—and particularly utilitarian—leanings. The question naturally arises as to whether and how the rule utilitarian can justify a non-maximizing action—whether this amounts to 'rule worship', as Smart (1956) charges. Sellars has a principled objection to this charge—one that he lays out in IIOR (§§80–83), and which in its essential details matches the solution offered in Rawls (1955).

This is well-tread ground, and I do not wish to recapitulate these debates. For I do not think that we have to read Sellars as a consistent utilitarian; nor do I think that consequentialism is the best way to make sense of morality as team reasoning. In Sellars's final (and arguably, most mature) writing on

ethical theory, all talk of maximizing has disappeared. Sellars states the 'regulative principle of the moral point of view' as follows:

> Shall [each of *us* do that which, in the circumstances, promotes the happiness of each and every one of *us*, all relevant things considered].
>
> (ORAV: §194)

This intention has as its 'core' a 'state of affairs intention… which is the formal specification of an intrinsic and encompassing good':

> Shall be [each and every one of us leads a satisfying life, all relevant things considered]
>
> (ORAV: §195)

I suggest this approach indicates a move away from Sellars's earlier utilitarianism, in that all talk of maximizing has disappeared (and in fact, the word appears not even once in the entire essay), and the explicit use of 'each and every' emphasizes the Kantian *separateness of persons* that is alien to the utilitarian outlook.

Our starting point lies in Sellars's claim that practical reasoning presupposes some substantive good.[7] Suppose we identify the following state of affairs as good:

> Shall be [each and every one of us tells the truth, all relevant things considered]

Suppose the following factual entailments obtains:

> (I tell the truth in circumstance C) implies (Smith and Jones will tell lies)

(I tell a lie in circumstance C) implies (Smith and Jones will tell the truth)

Does it follow that I ought to tell a lie in circumstance C? *Only if we think that intrinsic goods generate maximizing obligations rather than deontological obligations*. But there is no reason to think this—and (in light of various problems with consequentialist moral theory) many reasons to reject it. Thus, the door is opened to seeing morality in terms of a different kind of team reasoning (PTR) which is deontological rather than consequentialist or maximizing. So, to take Sellars's model in ORAV, the schema for such reasoning is (as indicated above in Section 5) as follows:

Shall be [each and every one of us leads a satisfying life, all relevant things considered]

'Each and every one of us leads a satisfying life, all relevant things considered' implies 'Each of us does A_1 in C_1, A_2 in C_2, etc'.

Therefore, Shall [Each of us does A_1 in C_1, A_2 in C_2, etc.]

Therefore, Shall_{we} [I do A_1 in C_1, A_2 in C_2, etc.]

Sellars's mature view is that the latter intention is a we-intention because it is derived from an intersubjective intention (Shall [Each of us does…]); he articulates this view in a number of places in his later works [ORAV: §178–180; Correspondence between Sellars and David Solomon (CSDS): §14].

A feature of PTR—one that Sellars insisted on throughout his career—is that one can act on an unshared moral intention. The following passage from ORAV is typical:

It is essential to note that the intention

Shall [*any* of us do *A*, if in *C*]

must not be confused with the intention

Shall [*everyone* of us do *A*, if in *C*].

The latter is an intention which is realized *only if* everyone of us who is in *C* does *A*. The former is realized *even if* only a single one of us who is in *C* does *A*.

(ORAV: §167)

This feature of moral reasoning is difficult to capture if we understand team reasoning in strictly consequentialist or maximizing terms.

Does it follow, then, that if my doing A_i in C_i follows from the intrinsically reasonable moral intention, that my doing A_i in C_i is reasonable for me, even if no one else shares the intention to do so? The answer, for Sellars, is 'it depends'. Sellars isn't Kant, and we should expect Sellars's deontology to allow for a certain amount of nuance regarding circumstances of action. 'Fiat justitia ruat caelum' might work for Kant, but is less appealing to the modern moral theorist. And indeed, Sellars does offer us such a flexible theory, although much of this flexibility is built into the circumstances clause. Consider, for example, his discussion in 'Reason and the Art of Living in Plato' (RAL), where he distinguishes the principles of a craft (such as housebuilding) from enforced imperatives (which may or may not coincide with these principles). The builder 'will regard this latter enforced imperative as throwing no light on what he ought to do *qua* builder, but as by no means irrelevant to what he ought to do *qua* having hungry mouths to feed' (RAL: §32). Adapting Sellars's discussion from RAL for the present context, suppose we have the following moral principle:

In C_k one ought to do A_m

This principle results from team-reasoning about promoting the general welfare, and the circumstance clause presupposes general compliance. But suppose we know that most others will not follow this principle. In that

case, general non-compliance becomes part of the circumstance. Thus, simplifying Sellars's discussion from RAL, we now have not C_i , but C_k (where "k" includes general non-compliance as part of the circumstances of action). It might be the case that

In C_i one ought to do A_m

but it might instead be the case that

In C_k one ought to do A_n

Thus, real-world circumstances (including ones that introduce immorality or evil into a situation) can alter our circumstances, and hence make different obligations binding on us. So, while principled team reasoning allows us to act on a moral principle without waiting for others to do the same, it is also not a foolish view that requires blind adherence to the same set of principles regardless of circumstances. Thus, while Sellars's deontological account of team reasoning will certainly admit of circumstances where unshared action is obligatory, it will also admit that circumstances counsel a different course of action. It is worth noting, however, that this different course of action might *still* be largely unshared. For example, if there is an enforced imperative such as 'Members of minority group M must be reported to the authorities for immediate arrest and deportation', you might find yourself under a moral obligation to lie to the authorities (and might do so), although this is a largely unshared action.

This all raises a number of thorny difficulties—but these are difficulties that are general to deontological accounts which admit of flexible or *prima facie* principles; I do not think that there is any problem here specific to Sellars's account. My point is that on Sellars's account of team reasoning, one cannot only have unshared commitments that nevertheless count as legitimate team commitments—one can also have unshared intentions that count as group intentions, and it can be reasonable to act on these intentions.

9 Conclusion

The development of theories of team reasoning has offered an exciting new tool to explain the moral point of view. I have argued, however, that this tool is limited if we insist on understanding team reasoning in consequentialist terms. If we do so, we seem constrained to understand morality as an assurance game. But I have argued that team reasoning itself need not have a consequentialist character, and that important features of morality can only be captured by a deontological species of team reasoning. By drawing on the work of Wilfrid Sellars, I have tried to elaborate a theory of Principled Team Reasoning that explains how morality involves reasoning from a collective point of view while still maintaining important central features of moral reasoning—in particular, the fact that morality may occasionally require us to behave in noncooperative ways. While obviously much more remains to be said, I hope this suggestion can offer a useful corrective to dominant team reasoning approaches to morality.

Notes

1. For an argument that practical reasoning involves reasoning among intentions, see Koons (2019), especially Chapter 3.⏎
2. If a person were truly epistemically isolated—raised in a cult-like environment, for example—we might be inclined to relativize, epistemically: We might say, for example, that certain of Hannah's belief, which would be unjustified if held by members of the larger community, might nevertheless be justified for her.⏎
3. For an argument that the coordinating function of our epistemic practice is in principle distinct and separable from its purely epistemic goals, see Koons (2025: Section 2.7.1).⏎
4. Tomasello largely appeals to mutualism as a partial *explanation* of the development of modern human morality

rather than as a *justification* for compliance with moral norms.⏎

5. The Organization case can be seen as an instance of agreeing on the rules, but disagreeing on what follows from them. The moral case is more often a case of disagreeing about what the rules are in the first place. But again, such disagreement occurs against a larger background of agreement over other rules, values, and commitments against which such disagreements may not be in principle unresolvable.⏎
6. I earlier denied this (Koons 2019: Ch. 4), and was willing to concede that such non-shared we-attitudes could be we-attitudes in the weak sense. Ronald Loeffler convinced me that this concession was a mistake.⏎
7. Here, Sellars anticipates later authors such as Hampton and Korsgaard, who argue that means-end reasoning generates no normative conclusions at all without a substantive conception of rationality telling us what is intrinsically valuable.⏎

References

Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Penguin Books.⏎

Griffin, Paul. 2017. *The Carbon Majors Database: CDP Carbon Majors Report 2017*. London: CDP. https://cdn.cdp.net/cdp-production/cms/reports/documents/000/002/327/original/Carbon-Majors-Report-2017.pdf (accessed on 3 October 2024).⏎

Koons, Jeremy Randel. 2019. *The Ethics of Wilfrid Sellars*. New York: Routledge.⏎

Koons, Jeremy Randel. 2022. ‘Knowledge as a Collective Status’, *Analytic Philosophy*, 63(4): 277–304.⏎
Koons, Jeremy Randel. 2025. ‘Commitment Coordination and the Social Function of Reason-Giving’, in Preston Stovall and Ladislav Koreň (eds), *Why and How We Give and Ask for Reasons*. Oxford: Oxford University Press, pp. 49–72.⏎
Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.⏎
Rawls, John. 1955. ‘Two Concepts of Rules’, *The Philosophical Review*, 64(1): 3–32.⏎
Searle, John. 2009. *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.⏎
Sellars, Wilfrid. 1967a (SM). *Science and Metaphysics: Variations on Kantian Themes*. The John Locke Lectures for 1965–6. Atascadero, CA: Ridgeview Publishing Company.
Sellars, Wilfrid. 1967b (OIM). ‘Obligation, Intersubjectivity and the Moral Point of View’, Chapter 7 (pp. 175–229) of *Science and Metaphysics: Variations on Kantian Themes*. The John Locke Lectures for 1965–6. Atascadero, CA: Ridgeview Publishing Company. Reprinted in MOP, pp. 138–187.
Sellars, Wilfrid. 1973 (RAL). ‘Reason and the Art of Living in Plato’, in Dale Riepe (ed.), *Phenomenology and Natural Existence: Essays in Honor of Marvin Farber*, pp. 353–377. New York: The University of New York Press. Reprinted in MOP, pp. 116–137.

Sellars, Wilfrid. 1976 (CSDS). Letter to David Solomon, 28 June 1976. Wilfrid S. Sellars Papers. Box 166, Folder 3, Archives of Scientific Philosophy, University of Pittsburgh Library System.

Sellars, Wilfrid. 1980 (ORAV). ‘On Reasoning about Values’, *American Philosophical Quarterly*, 17(2): 81–101. Reprinted in MOP, 465–510.

Sellars, Wilfrid. 2023 (MOP). *The Metaphysics of Practice: Writings on Action, Community, and Obligation*, edited by Kyle Ferguson and Jeremy Randel Koons. Oxford: Oxford University Press.

Smart, J.J.C. 1956. ‘Extreme and Restricted Utilitarianism’, *The Philosophical Quarterly*, 6(25): 344–354.⏎

Tomasello, Michael. 2009. *Why We Cooperate*. Cambridge, MA: The MIT Press.⏎

Tomasello, Michael. 2014. *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.⏎

Tuomela, Raimo. 2002. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press.⏎

Tuomela, Raimo. 2007. *The Philosophy of Sociality: The Shared Point of View*. Oxford: Oxford University Press.⏎

US Centers for Disease Control. 2024. ‘History of Smallpox’. https://www.cdc.gov/smallpox/history/history.html (accessed on 3 October 2024).⏎

Wilson, E.O. 2012. *The Social Conquest of Earth*. New York: Liveright.⏎

OceanofPDF.com

10 Social Reflective Equilibrium and Moral Progress

Sushruth Ravish

DOI: 10.4324/9781003612827-11

1 Introduction

Reflective Equilibrium (RE) is often described as the most defensible method we have for justifying moral claims (Scanlon 2002; DePaul 2013). While Nelson Goodman's work was an early precursor, RE as a method of theory construction and justification is most famously credited to John Rawls, with noteworthy later contributions by Norman Daniels (1979) and Thomas Scanlon (2014). Ever since it was brought to prominence it has not only interested meta-ethicists but also practitioners in bioethics, medical ethics, and population ethics as a viable method for building consensus and convergence on moral issues (Beauchamp and Childress 1994; Arras 2009; April and April 2016). In this chapter, I adopt a modest and local conception of moral progress, which involves improvements within particular communities on specific moral questions, even if such improvements are partial and do not extend across all domains (Moody-Adams 1999). The abolition of chattel slavery in North America is a paradigmatic example of such community-level progress. If RE is as central a method of moral justification as the literature suggests, one would expect it to help explain how these improvements arise. Indeed, some have argued that the growing acceptance of same-sex relations can be modeled through RE (Harris 2005). While decriminalization of homosexuality can, of course, be an example of desirable moral change, the case for it has been built largely on Anglo-American contexts. This chapter presents universal suffrage and the provision of minority rights in the Indian Constitution as

equally relevant examples worthy of discussion. This is not an effort to merely select a non-Western example for the sake of doing so. As I will attempt to show, in addition to the momentous nature of these proclamations, the debates during the drafting of the constitution (referred to as the Constituent Assembly Debates) provide as detailed documentation as is possible for any forum tasked with drafting a document providing ethical and legal guidelines to be followed by millions of individuals in a diverse country like India. Thus, I believe that there is much to be learned from the Indian experience, not only in terms of what the Constituent Assembly succeeded in doing but also where they fell short (judgments that are all too apparent now in hindsight). The analysis draws on work in social ontology and collective reasoning, to clarify how moral inquiry can be understood and evaluated at the level of collectives.

The first section of this chapter begins by clarifying the concept of moral progress employed here and the sense in which it could be modeled. The section then sketches the structural elements common to the method of RE and evaluates criticisms of RE. A modified version, called Social Reflective Equilibrium, is proposed to alleviate some of these concerns, demonstrating how communities can engage in collective reasoning and aspire to reach moral judgments that might be considered an improvement over those held by individuals earlier. The second section of this chapter will build on insights from the collective reasoning discourse to substantiate the claim that group-reasoning advantages identified in theoretical reasoning can be applied to the practical sphere, thereby refining our model in terms of its goals and responses to disagreements. The third section presents a case study, drawing on selected examples from the Indian Constituent Assembly Debates, to examine how the debates manifest (or fail to do so) the goals of collective reasoning and what lessons they hold for one another. In doing so, the chapter draws on and contributes to debates in social ontology and collective epistemology, which examine how groups reason, coordinate, and improve their shared moral judgments.

2 Reflective Equilibrium as a Model for Moral Progress

The last decade or so has seen a renewed philosophical interest in ideas related to moral progress (Buchanan and Powell 2018; Kitcher *et al*. 2021; Sauer *et al*. 2021). However, moral progress remains a contested concept, with disagreements over whether it should be understood broadly or narrowly, locally or globally, and over what criteria should be used to evaluate claims about moral progress. For the purposes of this chapter, I will follow the strategy of taking particular examples (like the abolition of chattel slavery and decriminalization of homosexuality in North America and Europe, and the prohibition of untouchability and protection of minority rights in India) that most people would agree mark a positive moral change in a community and attempt to explain how that change could have come about. As long as these particular examples are uncontroversial, the chapter will remain non-committal on what the explanations might entail for the concept of moral progress itself.[1] A point where I do take a side is on the issue of whether moral reasoning can play a role in bringing about moral progress. I attempt to demonstrate in Section 3 of this chapter that moral progress can be achieved when, through deliberation and reflection, people adopt new beliefs, reject old ones, or update their confidence in existing ones (Rehren and Blunden 2024). Among scholars who support the moral reasoning view, there have been attempts to model the process of belief transformation (Kumar and May forthcoming; Campbell and Kumar 2012). While I will not discuss the comparative merits and demerits of such attempts, the possibility of modeling progress through moral reasoning can align with the arguments presented here.

Given the widespread acceptance of RE as one of the most important methods of moral justification, it is worth undertaking an attempt to model moral progress through RE. Before we undertake this pursuit, however, a few preliminary issues must be addressed. Reflective equilibrium involves adjusting individual judgments, the principles we take to guide them, and the background considerations we deem relevant, so that they fit together in a coherent and mutually supporting way (Daniels 2023). However, multiple variants of RE are prevalent,[2] and not all are compatible with one another. In an effort to find the most amenable form of RE for this project, I begin by identifying the two essential structural elements that any plausible account of reflective equilibrium must include.[3] First, there are inputs that

we might call ‘considered judgments’, which are drawn upon while moral theorization is attempted. Considered Judgments are essentially, screened and filtered moral judgments from our initial stock of moral judgments. Such screening must occur under conditions ‘favorable for deliberation and judgment in general’ and ‘to the exercise of the sense of justice’ in the absence of conflict of interest and distortion (Rawls 1999: 42). Norman Daniels adds that any judgments that appear suspect upon reflection must be eliminated, and only those that can be asserted with confidence must be retained (Daniels 1980: 85–86). Thomas Scanlon views these considered judgments as moral judgments of any generality, including judgments about particular actions, moral principles, or any judgments about the kind of considerations that are of relevance to determining the rightness of actions (Scanlon 2014: 76–77). For instance, if one judges that caste discrimination is wrong and that equal political rights should not depend on social status, these judgments—when made under favorable conditions—can serve as considered judgments that later guide the selection of principles. Thus, the procedure for filtering and arriving at considered moral judgments is not to be seen as ad hoc but as a prescription for a general procedure for any considered judgments delivered under favorable conditions.

Once the inputs are filtered to yield considered judgments, operations are performed on these inputs to seek the maximum possible coherence among the inventory of considered judgments. To illustrate, from the considered judgment that caste discrimination is unjust, one might move to a more general principle forbidding forms of institutionalized exclusion. The interplay between the two levels clarifies how adjustment works in practice. As part of this stage, more general moral principles are to be formulated to systematize and explicate the considered judgments. These independently feasible principles are intended to be those that can account for considered judgments. It is expected that discrepancies, divergences, and conflicts will arise among the considered judgments and the general principles. Rawls suggests ‘dropping and revising… reformulating and expanding’ them till there is coherence or a ‘systematic organization’ between the two, which helps us reach a reflective equilibrium (1974: 8).[4] Such a reflective equilibrium, although an incremental improvement over considered moral judgments, is still ‘narrow’ and must not be seen as a destination (Rawls

1999). The circle of judgments can then be widened to include background theories (including non-moral judgments, such as the epistemic or pragmatic importance of various moral theories). Coherence is then sought between this enlarged set of judgments. A 'wide' reflective equilibrium would seek maximal coherence between '(a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories' (Daniels 1979: 258).[5]

Before we use such a quintessential model of RE to analyze how moral progress might be achieved, two significant concerns about RE need to be addressed.[6] First, the method of RE appears as a tedious and onerous task. The plausibility of such an exercise rests on an assumption of a rational self that employs a reasoning process that has impeccable discernment. What lends gravity to this reasoning process is that coherence-seeking involves not only finding deductive relationships between propositions but also ensuring that these relationships are consistent with one another. If coherence were a purely deductive matter, the method could have been easily formalized. However, reflective equilibrium involves exploring mutually supporting (and explanatorily useful) relationships between different elements, and therefore requires an involved and invested reasoning process that engages in the trade-offs between diverse elements. Any adulteration in the reasoning process can seriously jeopardize the hopes of achieving RE.

Increasing empirical evidence from cognitive science and moral psychology suggests this fear might be genuine. For instance, framing effects can be found when people's choices vary depending on how the question is worded or framed.[7] In the now-famous Wason selection test, which required an elementary understanding of *modus ponens* and *modus tollens*, only about 10% of the respondents gave the correct answer (Mercier and Sperber 2011: 63). Haidt *et al.* in fact, have coined the term 'moral dumbfounding' (MD) to refer to 'the stubborn and puzzled maintenance of a moral judgment without supporting reasons' (2000: 1). Jonathan Haidt argues that 'moral reasoning is an effortful process, engaged in after a moral judgment is made, in which a person searches for arguments that will support an already-made judgment' (2001: 818). Moral reasoning is presented as a post-hoc endeavor produced mainly to justify already-

formed moral judgments to others. As Haidt puts it, reasoning works more like a lawyer asked to defend a client than a scientist seeking the truth. If such psychological experiments are to be trusted and their conclusions granted, then, in situations where deliberate decision-making strategies are used, sub-optimal outcomes are produced because reason seems to function mainly as a self-justificatory apparatus rather than an objective, truth-finding inquiry. For our argument, even if there is a slight possibility that cognitive or implicit biases are prevalent, we should be wary of our chances of achieving moral progress through RE.

Second, traditional interpretations of RE have primarily advanced an individualistic understanding.[8] For example, Daniels suggests that RE aims to 'produce coherence in an ordered triple set of beliefs held by a particular person' (Daniels 1979: 258). Rawls, too, claims that,

> If we can characterise one (educated) person's sense of justice, we might have a good beginning toward a theory of justice. We may suppose that everyone has in *himself* the whole form of a moral conception.
>
> (Rawls 1999: 44)

Both Rawls' and Daniels' attempts, presented above, could be seen as paradigmatic of the dominant interpretation of RE as an internal, individualistic endeavor.[9] However, such an interpretation is not helpful when one is interested in understanding how communities accomplish moral progress. The paradigmatic example of moral progress cited in the literature is the abolition of slavery (Kitcher *et al.* 2021; Kumar and Campbell 2022), which was achieved through intense public deliberation and argumentation. It seems improbable that individuals working in isolation went about achieving RE for their respective belief sets and somehow converged on their opinions.[10] On the contrary, it is more defensible to claim that if not for widespread deliberations and argumentations, such reforms would not have been possible.

This chapter suggests that a modified account of RE, called Social RE, could better model moral progress than a more traditional interpretation.

Elsewhere, it has been argued that Social RE could serve as the epistemology for moral naturalists (Ravish and Sirola 2023). Considered judgments, when they are successful in theory reference, can provide reliable knowledge about moral properties because the reference of these moral terms is socially coordinated. Such a proposal amends traditional interpretations of RE, where considered judgments are individually held without reference to others and could not have laid any substantive claim to objectivity. Without rehashing metaethical debates about objectivity and moral realism, the aim here is to highlight how Social RE construes moral inquiry as a species of social inquiry that prioritizes interpersonal deliberations and argumentations. As individuals, we often have substantial limitations regarding the quantity and quality of life experiences we possess, our ability to process information, and how we empathize with others. However, as a group, through appropriate discussion and argumentation, we could overcome these lacunae. Interactions with others can broaden our horizons and help us identify blind spots in our evaluations. A dialectical engagement brings new considerations to our attention and can improve the quality of our deliberations and must be treated as an essential part of any moral inquiry. The argument is that, even with diverse starting points, the method of Social RE can resolve disagreements because it is primarily revisionary, with provisions to alter our starting presumptions if needed.

Social RE receives support from research in collective epistemology, which shows that groups perform better in conducive environments than individuals working in isolation. In various experiments involving mathematical or conceptual problems, it was found that groups showed a marked improvement in performance over when the problems were tackled individually. In many cases, the *assembly bonus effect* was observed, where even when no single person had the correct answer, the group collectively solved it through debate and discussion (Collins and Guetzkow 1964).[11] In a clear departure from the traditional individualistic approach, which views reason as an individual enterprise, Hugo Mercier and Dan Sperber contest the longstanding assumption in philosophy that solitary, contemplative reason is a means for individuals to acquire superior knowledge and make better decisions (2017). Taking an adaptationist approach to understanding

reason, Mercier and Sperber argue that at its heart, reasoning is a ‘social competence’ that ‘evolved as a response to problems encountered in social interaction’ (2017: 182–183). Through their social interactionist theory of reason, they provide a convincing account explaining the apparent incongruity: why reason behaves as a self-justifying tool that contributes to poor epistemic outcomes while excelling in a more social role. Their account of reason does not conflict with experimental evidence but builds upon it, making predictions about the output it is most likely to produce. They contend that the social nature of reason can be understood in terms of how it aids coordination by producing justifications and communication by producing arguments. The reason involved in the production of arguments concerns itself with achieving a high level of satisfaction, seeking arguments that are just good enough for those involved, without being the best.

Additionally, when individuals engage in arguments, they often display confirmation bias. While confounding the rival individualistic views, these features appear to be in place for the interactive view proposed by Mercier and Sperber. Their argumentative theory of reasoning acknowledges and accounts for the mere rationalizing role that reason plays in an individual reflective mode. As highlighted by Haidt and others, reason’s suboptimal performance does not appear to be a fatal flaw, but rather a failure to understand its function. Reason accrues epistemic benefits, making communication reliable and effective. However, this can occur only when reason is deployed in a social deliberative setting. Mercier and Sperber argue that this social interactive reasoning ensures, in cognitive tasks, for instance, the success of a group, even while individuals working in isolation might have failed. Their account, when adopted, recognizes and accommodates the intrinsically social nature of moral inquiries.

For any criticism that expresses skepticism about the efficacy of group deliberations on moral issues, Elizabeth Anderson’s works could offer a response. Anderson points out that even when we lack an accurate understanding of moral goodness, we may still possess knowledge of moral errors and prejudices. Learnings from identifying the factors and conditions that contribute to veritistic achievements and decrease errors in non-moral issues can also be applied to moral problems with some confidence.

> While we do not already have on hand a final standard of moral rightness, we may have a fairly good idea of characteristic sources of moral error, ignorance, bias, and blindness drawn from social and cognitive psychology and historical investigation.
>
> (Anderson 2016: 93)

Anderson thus lays a clear possibility of a naturalized social moral epistemology, even if we were unsure of the ontology and metaphysics surrounding moral truths. It is worth noting that the claim is not that collectives consistently outperform individuals in arriving at correct judgments under all conditions. There could indeed be suboptimal settings for a group under which the results tend to be veritistically inferior. For instance, if all group members initially share the same opinion, then debates tend to stall. Homogeneity of the interlocutors' arguments could also lead to greater polarization of beliefs. In this regard, Anderson suggests that a democratic, social moral epistemology is necessary to ensure that moral progress is achieved by groups (Anderson 2016: 78).[12] Thus, moral progress is more likely to be achieved under participatory social arrangements that help us overcome biases and prejudices and improve access to information. In their absence, an authoritarian social arrangement that relies heavily on reflections of only a few privileged individuals is unlikely to yield moral progress. While reflection can fetch us moral insights, it is unlikely to emanate solely from the expression of someone’s wisdom. As a society, whenever we claim moral progress, it is more likely to have emerged from intense public argumentation, as history shows through reformative policy decisions in various societies.

3 Collective Reasoning and the Possibility of Moral Progress

The account of Social Reflective Equilibrium proposed here essentially serves as a model of collective reasoning grounded in deliberation. This

brings any researcher interested in Social RE into conversation with social-ontological accounts of how groups can function as sites of shared judgment and reasoning. The endeavor in this section is to build on the work in collective reasoning done by Christopher McMahon (2000, 2001) to substantiate further the claim that the results found in theoretical reasoning can carry forward to the practical sphere. I attempt to delineate Social RE so as to fill in details about (a) how social RE is supposed to be put into place beyond saying it would involve an exchange of reasons and mutual adjustment of judgments and principles and (b) who gets to be a part of the collective; whether some members would be considered to be epistemically advantaged in this dialectical process and if there should be certain pre-conditions put into place to ensure that the process kicks off and fulfills its mandate. McMahon explicitly suggests that he is 'using the method of wide reflective equilibrium' to answer the question of 'exactly how the requirement to contribute is to be understood' (2000: 322). Therefore, his work on collective rationality and collective reasoning provides a relevant extension of Social RE. In addition, finer details could be worked out regarding the goals of reflective equilibrium and how it could proceed, building on the work of scholars studying collective moral progress.

This chapter does not attempt to provide a comprehensive account of the collective reasoning McMahon proposes. Instead, it will strive to delineate those aspects that have a direct bearing on the activities that constitute the process of Social RE. McMahon argues against a standard aggregative model in which preferences are aggregated into a social decision. This is especially relevant since many critics have pointed out how the standard 'utility-maximizing conception of practical rationality' promotes defection from cooperative schemes to further one's private interests. McMahon agrees with thinkers like Joshua Cohen, who hold that deliberative democracy stands for an 'association whose affairs are governed by the public deliberation of its members' (as quoted in McMahon 2001: 87). However, McMahon cautions us against interpreting deliberative democracy as being constituted solely by the public deliberation of its members, as such efforts could only be carried out by numerically small collectives and for only short durations. Instead, we should understand deliberative democracy to be an association where each member gets an

equal opportunity to influence others, such that even if the group had to vote, in the absence of a consensus, there would be an appeal to public reason, such that members could rise above their competing claims and attempt a resolution. McMahon advocates deploying the Rawlsian notion of public reason, which consists of a 'set of principles and values that are appropriate for guiding the common affairs of a pluralistic society' (2001: 90). These public reasons are grounded in reasonableness in the sense of what could be reasonably accepted by the members of a pluralistic society. The claim is that,

> When voting is preceded by shared deliberation about what is substantively reasonable-as-fair, and the members of the group express their opinion about this in their votes, the policy decided upon is more likely actually to be substantively reasonable-as-fair than it would be in the absence of deliberation.
>
> (McMahon 2001: 100)

In contrast to aggregative, non-deliberative collectives, deliberative processes aim to transform the preferences of those who find themselves in the minority, such that relevant reasons motivating the group's decisions are tracked through deliberation (McMahon 2001: 95). When such a radical transformation occurs, a consensus emerges. However, if disagreement persists after deliberation, claiming superiority over a standard aggregative model might be challenging. McMahon argues that epistemic virtues are secured by the deliberative aspect of decision-making (2001: 96). As long as the group wishes to track important reasons for taking action, it shouldn't be viewed as a defeat for those who do not succeed. They can believe that those who do succeed are serving the entire group by identifying the right course of action (McMahon 2001: 96). Even when decisions have moral dimensions, it might help the minority reconcile their moral concerns as long as democratic results are achieved. One obvious way shared deliberations might help in moral inquiries is by enabling interlocutors to improve their understanding of the moral requirements through active

participation in the process. As mentioned earlier, the assumption is that dialectical exchanges might convert the members of the collective to the correct view. Two further effects of shared deliberations have been highlighted. First, the minority can take heart from any changes in the majority's viewpoint brought about by the minority's criticisms. This is because such changes could be taken to reflect an improved understanding of the relevant values from their standpoint. Second, when interlocutors face criticisms for their views but find such criticisms plausible, it might increase confidence in the others' intellectual abilities. The overall hope is that, while there might be significant deviations between the correct view and the view the group initially tends toward, the degree of deviation will gradually decrease (McMahon 2001: 100–102).

A more fleshed-out understanding of Social RE can also be obtained if we pay attention to how McMahon defines collective reasoning and its goals. McMahon defines collective reasoning as 'the cooperative assessment of the rational force of the putative facts offered as reasons' (2001: 104).[13] The fact that the individuals (part of the process) are cooperating is already assumed. Cooperation is assumed because participants are assured something of value as a result of the cooperation—namely, 'a justified judgment' that is adequately supported by reasons and one that is supposed to be epistemically superior to what the interlocutors would have accomplished on their own. This judgment identifies and evaluates the relative significance of the reasons pertinent to the question (McMahon 2001: 105). McMahon's conceptualization aligns well with RE because RE (at least according to Rawls) is agnostic about the question of truth in the inquiry.[14] Justified judgment is the goal of the inquiry, and this would be agreeable to scholars favoring warranted assertions in place of truth; scholars partisan to truth could concede that justification is the route to truth and hence be willing participants.[15] McMahon argues that the cooperation involved in collective reasoning should go beyond mere aggregation of reasons into creating 'a common pool of criticized arguments germane to the issue the group is addressing' (McMahon 2001: 109). Each group member is responsible for drawing on this common resource and arriving at the requisite judgments. The outcome of such an activity would be what McMahon calls a 'piecemeal consensus', which is a

‘de facto concurrence of judgments that can legitimately be made independently, and thus could be made by the members of a group of investigators at a different time’ (2001: 112).

McMahon thus stops short of designating ‘collective judgment’ as an outcome of collective reasoning. Collective judgment is a much more stringent requirement according to which each member would suspend judgment about a particular conclusion until all members can claim to have reached the same conclusion. The corresponding outcome would be an integral consensus where ‘acceptance by all is a condition for acceptance by anyone’ (McMahon 2001: 112). While collective judgment appears to take into account interpersonal reasons, it also carries a risk of debilitating and stalling the outcome of the collective. McMahon’s releasing collective reasoning from the burden of collective judgment or creating an integral consensus also works well for Social RE. If Social RE can be taken to have only piecemeal consensus as its goal, the familiar worries over the impossibility of agreement among the interlocutors could be warded off.

McMahon also provides resources to address traditional criticisms of RE regarding the prevalence of disagreement. McMahon’s contribution is especially relevant because, in the discussions over the application of Social RE in applied ethics, there is rarely a mention of how to deal with interpersonal disagreement. Paralleling existing discussions in the literature on moral disagreement, McMahon points out that disagreement can arise either with or without cognitive malfunction among members of the collective. The former case of cognitive malfunction could be further analyzed into cases where members might be mistaken about the implications of relevant reasons. This could be due either to not possessing the relevant concepts or to an error in applying the concepts one possesses. In the latter case of disagreement without cognitive malfunction, the cause may be inherent variability in the application of concepts or the use of an equally legitimate but alternative set of concepts. In many moral inquiries, this latter case manifests. Disagreement could persist because of the ‘burdens of judgment’ involved in public reasoning, a domain where there is an inevitable variability in the application of concepts. Building on Gutmann and Thompson (1996), McMahon suggests an ‘economizing of moral disagreement’ wherein one should choose those reasons for one’s position that most accommodate the views of those opposing us. While this

is acknowledged not to eliminate disagreements, it does provide a standard for what constitutes an appropriate resolution of disagreements when there are conflicting moral commitments.

This section began with two questions regarding the implementation of Social RE. The discussion of collective rationality based on McMahon's works has addressed the first question of 'how' or the mechanisms by which Social RE would be implemented. I would now like to address the second question of 'who' or the issues related to participation in the collective deliberation process. To begin with, we would do well to heed Anderson's worries about how, in collectives, both individuals and dominant social groups can indulge in authoritarianism and be prone to moral blindness (Anderson 2016). Moral inquiries become authoritarian when conducted by those who are privileged and benefit from being perched atop social hierarchies. Those at the receiving end of these hierarchies are either excluded from participation or their responses do not get the requisite uptake (Anderson 2016: 78). Using case studies from the abolitionist movement in Europe and North America, Anderson convincingly demonstrates how privilege and a lack of relevant exposure to lived experiences contributed to apparent moral blindness among otherwise well-meaning individuals (2016: 83–94). While it would be evident to any outside observer, the presence of self-serving bias and other epistemic vices caused the concerned individuals to be oblivious to the contradictions between their proffered ideals and their treatment of Black people.

As a remedy for the perils of moral blindness and authoritarian tendencies of the powerful groups, Anderson argues for a democratic moral inquiry. Like McMahon, Anderson insists that here, 'democratic' should not mean just an outcome determined by majority voting, but rather equal participation of members, where there would be genuine 'receptiveness to others in their full humanity' (2016: 94). Such receptiveness would also involve acknowledging that, in many instances, the oppressed have an epistemic advantage in perceiving the violation of moral principles more clearly than the dominant.[16] This would also be backed by debates in standpoint epistemology, where it is argued that the marginalized are in a better position to acquire moral knowledge. The oppressed, in contrast to the privileged, are better at interpreting evidence, distinguishing between

relevant and irrelevant moral considerations, weighing the moral significance of issues, and determining their competence in applying concepts that bear on moral issues (Dular 2024). Anderson claims that in cases where collectives make significant progress, contentious politics can foster genuine deliberation and more effectively convey the claims of the marginalized.

> First, contentious politics—active, practical, mass resistance to the moral claims embodied in social institutions enforced by and catering to the powerful—is needed to activate genuine practical reasoning across all levels of society. The powerful won't really listen to reason—that is, to claims from below—until they no longer have the power to routinely enforce their desires. Second, the subordinated and oppressed must actively participate in that contention. They must manifest in deed and not only words their own interest, capacity, and worthiness for the rights and privileges they are demanding.
>
> (Anderson 2015: 39)

Thus, Anderson's account of moral progress is built on insights from non-ideal epistemology and advances a social moral epistemology that acknowledges the need for a collective and emphasizes the importance of securing the collective's constitution in ways that provide appropriate representation to the marginalized. Another deliberative account of reasoning that can complement Anderson's reasoning-based account of moral progress is a communitarian account of collective reasoning proposed by Agnes Tam (2020). Tam argues that social norms often trump moral norms and that when social movements do manage to accomplish moral progress in a society, it is because the members of the social group were able to provide convincing 'We-reasons'. For instance, the activists in the anti-slavery movement in Britain appealed to patriotism and argued that the inhumanity of slavery violated the honor of being British (Tam 2020: 95). She thus makes a case that for moral progress to occur, We-reasoning must

work in tandem with democratic reasoning. As she puts it, '[t]o improve our collective thought and action, the powerful need to learn from the oppressed. But the moral lessons must also infiltrate into the social grammar of communal life' (Tam 2020: 96). Thus far, we have seen how Social RE could act as a model for moral progress and how that would translate into a collective reasoning endeavor. In the last section, I will attempt to draw on a couple of examples from the Indian Constituent Assembly debates to highlight how the discussions exemplify or fail to achieve the goals and ideals of Social RE and collective reasoning.

4 A Case Study: Indian Constituent Assembly Debates

India's Constituent Assembly (1946–1950) was formed with the mandate of addressing not only issues of governance, such as the separation of powers and the structure of the government, but also to formulate a code of ethics that could be followed by all citizens of India (Bhargava 2022). The Constituent Assembly (henceforth CA) was formed in India through elections in July 1946, with each province allotted seats in proportion to its population. Separate electorates were established for Muslims, Sikhs, and the general communities. While the Muslim League was the second-largest party after Congress in 1946, with 73 seats to Congress's 202, Congress had 82% control after partition, with Muslim League seats falling to 28 (Austin 1972: 9–10). The princely states were allotted 93 representatives through nominations and elections. The minorities (Scheduled Castes and Tribes, Parsis, Anglo-Indians, and Nepalis) were brought into the assembly through Congress's suggestions, making up 37% of the provincial membership (as quoted in Bajpai 2016: 45). Granville Austin, in his landmark work on the making of the Constitution, remarks that the Congress leadership ensured that 'persons of exceptional ability' were present in the Constituent Assembly (1972: 11). This included many who were not from Congress at all (most notably B.R. Ambedkar). Austin also provides us with testimony from K. Santhanam, a member of the CA, observing that 'there was hardly any shade of public opinion not represented in the Assembly' (1972: 13). The CA was thus, 'a diverse collection of individuals and groups [that]

became the people of a single book, one that reflects their commitment to protect their mutual rights and which articulates a collective identity' (Bhargava 2008: 1). Members of the CA deliberated and debated for more than three years on the fundamental principles that guided the new republic. In the words of Rajeev Bhargava,

> The Constituent Assembly Debates demonstrate a common commitment to the idea that political outcomes will not be only a reflection of the self-interest of currently well-organised groups but will also embody inclusive principles which encompass even the interests of groups that are ill-organised or invisible at that time. This open-endedness indicates the willingness of people to modify their existing preferences—in short, to justify outcomes by reference not to self-interest but to *reason*. It also shows a willingness to recognise the creative value in difference and disagreement [emphasis added].
>
> (Bhargava 2008: 25)

Through examples, I will attempt to show that wherever 'reason' prevailed, both democratic reasoning and We-reasoning (advocated by Agnes Tam) worked in tandem to yield what we might call significant accomplishments of the CA. In particular, this section aims to draw the reader's attention to a genuinely noteworthy achievement—the adoption of universal suffrage.

While other countries initially granted suffrage to educated men who owned property, with race often playing a significant role, India was unique in offering universal adult franchise to all its citizens. Universal suffrage was intended to be part of fundamental rights; however, it was dispersed and found scattered in various articles, including Articles 66 and 149 on elections, and Article 289 on the Election Commission (Jha 2004: 4358). There was opposition to this proposal, with several members, including Brajeshwar Prasad, M. Thirumala Rao, K.T. Shah, and Hridya Nath Kunzru, expressing varying levels of resistance (Constituent Assembly

Debates 11.162.211; 11.162.74). Most of these speakers were worried that the masses in India were not educated enough to handle this responsibility. They felt that large chunks of the population might not realize the value of patriotism and would be motivated by communal and regional agendas. In response to these criticisms, Krishna Chandra Sharma argued that a universal adult franchise was essential to uphold the sovereignty of the people and express the will of the nation as a whole (Constituent Assembly Debates 11.159.132). Alladi Krishnaswami Ayyar, too, argued that anything stopping short of universal adult suffrage would negate the principles of democracy. He felt that the CA had taken this necessary step with faith in the common people, and the collective should be congratulated for that (Constituent Assembly Debates 11.163.60). One must agree with Gautam Bhatia when he remarks, 'Here, the Constitution was transformative: every adult Indian was a citizen, presumed to be autonomous and responsible enough to be capable of participation in the core public activity of democracy' (Bhatia 2018: 120). Thus, although there was some initial resistance to the idea, members ultimately appeared to rally behind the issue. Ultimately, as a collective, CA achieved what has been hailed as a remarkable milestone.

We should take note of a dissenting opinion amidst the self-congratulatory tone of the CA. Kamlapati Tiwari rhetorically asks the assembly, 'If this great democracy, which you are going to establish, is not based on the rights of the people, that is to say, on adult franchise, on what else will it be based?' (Constituent Assembly Debates 11.163.162; 11.162.163). He claimed that the adult franchise was already presupposed in the freedom struggle and in our proclamation of democracy. Indeed, the cry for adult franchise can be traced back to the Constitution of India Bill of 1895. The demand appears almost inextricably linked to the freedom movement. The Nehru Report of 1928, often considered a pertinent document of reference for CA, too, made explicit calls for universal adult franchise.[17] The argument then could be that the consensus agreement over the issue was not so much one motivated by reason, as much as it was motivated by a joint social commitment that all members felt towards their country. The obligations they owed to the spirit of the freedom movement

might have been as much a driving force behind the CA's acceptance of universal adult franchise as were the rational arguments of its proponents.

In contrast to the deliberative success of the universal adult franchise, researchers have also demonstrated how the overarching theme of national unity provided a normative vocabulary for debates and went on to influence various issues raised (Bajpai 2016: 70–115). For instance, women accounted for only 5% of the CA's strength, and no women served on the final drafting committee. When women's issues were raised, however, they were always viewed through the lens of nationalism, national unity, and Indian tradition (Devenish 2019). Insofar as the members gave these reasons to themselves and, by virtue of their social membership, they fit the description of We-reasoning argued for by Agnes Tam. It appears then that 'national unity' or 'the spirit of freedom movement/democracy' provided the members of the CA, We-reasons that were at least as influential as deliberative reasons. As pointed out by Alisha Dhingra,

> The demand for rights for women was linked to the eradication of immoral practices to enable women to fulfil their duties towards the nation. The emphasis on "nation" and women's duties towards the nation emphasised women's duties rather than on their rights.
>
> (Dhingra 2022: 40)

While more evidence and theorization are needed to make the case, these observations at least raise awareness of another possible criticism of We-reasoning. While Tam acknowledges the curtailment of participation as a concern for We-reasoning (Tam 2020: 93), what we find in our case study is not so much a curtailment of participation, but rather a subduing of women's issues under the guise of building a democratic nation-state (Devenish 2019). Discussions on issues like prostitution showed an overwhelming concern by all members for the morality of women, a crucial cog in the construction of Indian womanhood (De 2018). Interestingly, the women members who spoke also seemed to employ 'the discourse of "true womanhood" and "domesticity", which got entrenched during the

nationalist period' (Dhingra 2022: 40). While this could have been a strategic move to garner broader support, a possibility remains that the underrepresented chose to internalize the We-reasons provided by the grand projects of national unity and nation-building.

Lest any reader conclude that the CA debates served as fertile ground for communicative action with reasoned arguments and consensus-building, I bring to their attention instances in which 'bargaining' took place. As experts on constitutional debates define it, bargaining is a speech act that involves coercion and threats that fall outside the domain of deliberation (Bhatia 2018: 3). Given the enormity of the task the CA had in front of them and the diverse nature of its members, it is perhaps unsurprising that the transcripts of CA debates show instances of both reasoned debate and bargaining. Muslim leaders were heckled when they tried to propose innocuous amendments, which, for all purposes, were in the interest of strengthening democratic proposals (Nigam 2008: 129). Similarly, Rama Kant Agnihotri points out that while the issue of language was discussed with remarkable seriousness and integrity, '[d]iscussions would often get coloured by myths, prejudices, and stereotypes of India. These decisions were informed by monolithic and homogeneous views… in which the voices of minorities and the underprivileged had to be pushed under the rug' (Agnihotri 2015: 48). Thus, the painstaking detail with which the CA debates have been documented not only helps us see the radical and transformative nature of the constitution but also suggests perhaps where the makers fell short, or so it appears to us in hindsight.

5 Conclusion

Our moral judgments are rarely arrived at by solitary reasoning, and the social task of asking for and producing reasons and arguments must inextricably be part of moral inquiries if they are to achieve veritistic goals. While traditional interpretations of RE rely exclusively on individual rationality, this chapter proposes viewing moral inquiry as a community-wide enterprise that can eliminate subconscious biases and yield more morally justified consensus judgments. Recognizing that substantive moral progress can be achieved only through intense public deliberation led to the

investigation of works on collective reasoning and it was found that works by McMahon, Anderson, and Tam have much to offer in terms of the nuts and bolts of implementing a Social RE. Whether it is in the constitution of the collective that desires to adopt the method of Social RE or the mechanisms by which collective deliberations ought to proceed, valuable lessons are to be learned for anyone interested in moral progress for collectives. The point of using the Indian CA debates as a case study was not only to highlight instances of moral progress that a collective achieved, but also to learn from the shortcomings of that collective. Much remains to be uncovered about what CA debates have to offer in this regard. I hope that philosophers and researchers interested in collective reasoning will take this as an invitation to examine more closely how CA debates reveal underexplored aspects of social epistemology and social ontology. No community can take moral progress for granted, and a study of collective rationality and reasoning will serve us well if we want to understand how and when social inquiries can facilitate moral progress.

Notes

1. This is in line with the approach followed in Sauer (2019) and Buchanan and Powell (2018).⏎
2. A recent work by Kauppinen and Hirvelä (forthcoming) counts 256 possible variants.⏎
3. McPherson (2015: 655) has offered a similar analysis.⏎
4. Rawls does not adequately clarify what achieving coherence entails. It is often suggested that it involves maximizing relations of support and minimizing conflicts between the members of the belief set (Cath 2016). Consistency, systematicity, generality, and simplicity have all been considered as desirable features of the resulting belief set Kappel (2006).⏎
5. Background theories would include aligned non-moral judgments, such as the epistemic or pragmatic importance of

various moral theories.⏎

6. For the sake of brevity, the discussion of RE has been restricted to these two concerns. A more comprehensive defense of RE is attempted in Ravish and Sirola (2023), from where parts of Section 10.2 have been taken.⏎
7. When presented with logically equivalent instances of the ‘pigeonhole problem’, only 30% of the respondents gave a correct answer. But when provided a real-world context, around 70% gave the right answer. Thus, as the frames of these individuals varied, the importance they gave to reasons was found to vary even in the case of structurally fundamental problems.⏎
8. Barring some notable exceptions like Baderin (2017).⏎
9. To be fair, in one of his earlier writings, Rawls uses the first person plural ‘we’ while referring to how we can arrive at considered moral judgments and claims that the legitimacy of considered moral judgments is to be granted when competent judges arrive at a consensus (Rawls 1951: 180). However, he quickly abandons that in his later works, as seen in the quote.⏎
10. The issue resembles challenges raised in social ontology about how collective intentions are formed and sustained. Individualist accounts often struggle to explain such coordination (Meijers 2003).⏎
11. Some help can be taken from the results of the Condorcet Jury theorem, which shows that as long as members of a collective are more likely to be right than wrong about the answers to a particular question, then as long as the group is large enough, as a whole, the group mathematically tends to get the answer right.⏎
12. Anderson clarifies that by democratic, she intends equal and dignified participation of all stakeholders in a moral dispute (Anderson 2016: 94).⏎

13. Collective reasoning has also been defined as a ‘cooperative effort, involving linguistic exchange, to answer a question or solve a problem confronting a group’ (McMahon 2001: 105). It is this goal-driven nature of collective reasoning that distinguishes it from mere conversations, which might not amount to anything more than an exchange of observations.⏎
14. ‘The procedure of reflective equilibrium does not assume that there is one correct moral conception. It is, if you wish, a kind of psychology and *does not presuppose the existence of objective moral truths*. [emphasis added] (Rawls 1974: 9).⏎
15. Rawls himself was not in favor of a realist construal of RE. Rawls claims that the ‘history of moral philosophy shows that the notion of moral truths is problematical’ and asserts that it should be understood as a descriptive procedure (Rawls 1974: 9).⏎
16. Anderson gives the example of how it was the former colonies of Europe that took the lead in the legal application of the Universal Declaration of Human Rights, while countries like France, Britain, and the USA bickered over inserting a colonial clause that prevented subjects of colonies from claiming their rights.⏎
17. Available at https://www.constitutionofindia.net/historical-constitution/nehru-report-motilal-nehru1928/⏎

References

Agnihotri, Rama Kant. 2015. ‘Constituent Assembly Debates on Language’, *Economic and Political Weekly*, 50 (8): 47–56.⏎

Anderson, Elizabeth. 2015. ‘Moral Bias and Corrective Practices: A Pragmatist Perspective’, *Proceedings and*

Addresses of the American Philosophical Association, 89: 21–47.⏎

Anderson, Elizabeth. 2016. ‘The Social Epistemology of Morality’, in Michael Brady and Miranda Fricker (eds), *The Epistemic Life of Groups*, pp. 75–94. Oxford: Oxford University Press.⏎

April, Carolyn W., and April, Michael D. 2016. ‘Understanding Moral Distress Through the Lens of Social Reflective Equilibrium’, *The American Journal of Bioethics*, 16 (12): 25–27.⏎

Arras, John. 2009. ‘The Way We Reason Now: Reflective Equilibrium in Bioethics’, in Bonnie Steinbock (ed), *The Oxford Handbook of Bioethics*, pp. 46–71. Oxford: Oxford University Press.⏎

Austin, Granville. 1972. *The Indian Constitution: Cornerstone of a Nation, Indian edition*. Bombay: Oxford University Press.⏎

Baderin, Alice. 2017. ‘Reflective Equilibrium: Individual or Public?’, *Social Theory and Practice*, 43 (1): 1–28.⏎

Bajpai, Rochana. 2016. *Debating Difference: Group Rights and Liberal Democracy in India*. Oxford, New York: Oxford University Press.⏎

Beauchamp, Tom, and Childress, James. 1994. *Principles of Biomedical Ethics*. Oxford: Oxford University Press.⏎

Bhargava, Rajeev, ed. 2008. *Politics and Ethics of Indian Constitution*. New Delhi: Oxford University Press.⏎

Bhargava, Rajeev. 2022. *Between Hope and Despair: 100 Ethical reflections on Contemporary India*. New Delhi:

Bloomsbury Publishing.⏎
Bhatia, Udit, ed. 2018. *The Indian Constituent Assembly: Deliberations on Democracy*. South Asia edn. London: Routledge, Taylor & Francis Group.⏎
Buchanan, Allen, and Powell, Russell. 2018. *The Evolution Of Moral Progress: A Biocultural Theory*. Oxford: Oxford University Press.⏎
Campbell, Richmond, and Victor Kumar. 2012. 'Moral Reasoning on the Ground', *Ethics*, 122 (2): 273–312. https://doi.org/10.1086/663980.⏎
Cath, Yuri. 2016. 'Reflective Equilibrium', in Herman Cappelen, Tamar Gendler, and John Hawthorne (eds), *The Oxford Handbook of Philosophical Methodology*, pp. 213–230. Oxford: Oxford University Press.⏎
Collins, B. E., and Guetzkow, H. 1964. *A Social Psychology of Group Processes for Decision Making*. New York: Wiley.⏎
Constituent Assembly Debates, 1946–1950, The official report, Delhi University Library System, New Delhi; also at Indian National Archives, New Delhi; also see, https://www.constitutionofindia.net/constituent-assembly-debate/volume-11/.⏎
Daniels, Norman. 1979. 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *The Journal of Philosophy*, 76 (5): 256–282.⏎
Daniels, Norman. 1980. 'Reflective Equilibrium and Archimedean Points', *Canadian Journal of Philosophy*, 10 (1): 83–103.⏎

Daniels, Norman. 2023. ‘Reflective Equilibrium’, in Edward N. Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Fall 2023 edition). https://plato.stanford.edu/archives/fall2023/entries/reflective-equilibrium/.⏎

De, Rohit. 2018. *A People’s Constitution: The Everyday Life of Law in the Indian Republic*. Princeton, NJ: Princeton University Press.⏎

DePaul, Michael. 2013. ‘Reflective Equilibrium’, in H. Lafollette (ed), *The International Encyclopedia of Ethics*, pp. 4466–4475. Malden, MA: Wiley-Blackwell.⏎

Devenish, Annie. 2019. *Debating Women’s Citizenship in India 1930–1960*. New Delhi: Bloomsbury Publishing.⏎

Dhingra, Alisha. 2022. ‘Gender Discourses and the Making of the Indian Constitution’, *Indian Journal of Gender Studies*, 29 (1): 33–54. https://doi.org/10.1177/09715215211056794.⏎

Dular, Nicole. 2024. ‘Standpoint Moral Epistemology: The Epistemic Advantage Thesis’, *Philosophical Studies*, 181 (8): 1813–1835. https://doi.org/10.1007/s11098-023-02072-9. ⏎

Gutmann, Amy, and Thompson, Dennis F. 1996. *Democracy and Disagreement*. Cambridge, MA: Harvard University Press.⏎

Haidt, Jonathan., Bjorklund, Fredrik., and Murphy, Scott. 2000. ‘Moral Dumbfounding: When Intuition Finds No Reason’. Unpublished manuscript, University of Virginia.⏎

Haidt Jonathan. 2001. ‘The emotional dog and its rational tail: a social intuitionist approach to moral judgment’,

Psychological Review, 108 (4): 814-34. https://doi.org/10.1037/0033-295x.108.4.814.⏎

Harris, Charles Jr. 2005. 'Reflective Equilibrium as a Theory of Moral Change', *Southwest Philosophy Review*, 21 (2): 67–82.

Jha, Shefali. 2004. 'Representation and Its Epiphanies: A Reading of Constituent Assembly Debates', *Economic and Political Weekly*, 39 (39): 4357–60.⏎

Kappel, Klemens. 2006. 'The Meta-justification of Reflective Equilibrium', *Ethical Theory and Moral Practice*, 9: 131–47. https://doi.org/10.1007/s10677-005-9006-2.⏎

Kauppinen, Antti, and Hirvelä, Jaakko. Forthcoming. 'Reflective Equilibrium', in David Copp, Tina Rulli, and Connie Rosati (eds), *The Oxford Handbook of Normative Ethics*. Oxford University Press.⏎

Kitcher, Philip, Heilinger, Jan-Christoph, Jaeggi, Rahel, and Susan Neiman. 2021. *Moral Progress*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197549155.001.0001.⏎

Kumar, Victor, and May, Joshua. Forthcoming. 'Moral Reasoning and Moral Progress', in David Copp and Connie Rosati (eds), *Oxford Handbook of Metaethics*. Oxford: Oxford University Press.⏎

Kumar, Victor, and Campbell, Richmond. 2022. *A Better Ape: The Evolution of the Moral Mind and How It Made Us Human*. Oxford: Oxford University Press.⏎

McMahon, Christopher. 2000. 'Collective Rationality', *Philosophical Studies*, 98 (3): 321–344.

https://doi.org/10.1023/A:1018657013537.⏎
McMahon, Christopher. 2001. *Collective Rationality and Collective Reasoning*. Cambridge: Cambridge University Press.⏎
McPherson, Tristram. 2015. 'The Methodological Irrelevance of Reflective Equilibrium', in Chris Daly (ed), *The Palgrave Handbook of Philosophical Methods*, pp. 652–674. London: Palgrave Macmillan.⏎
Meijers, Anthonie. 2003. 'Why Accept Collective Beliefs?: Reply to Gilbert', *ProtoSociology* 18: 377–388. https://doi.org/10.5840/protosociology200318/1915.⏎
Mercier, Hugo, and Sperber, Dan. 2011. 'Why Do Humans Reason? Arguments for an Argumentative Theory', *Behavioral and Brain Sciences*, 34 (2): 57–74.⏎
Mercier, Hugo, and Sperber, Dan. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.⏎
Moody-Adams, Michele M. 1999. 'The Idea of Moral Progress', *Metaphilosophy*, 30 (3):168–185. https://doi.org/10.1111/1467-9973.00120.⏎
Nigam, Aditya. 2008. 'A Text Without Author Locating the Constituent Assembly as Event', in Rajeev Bhargava (ed), *The Politics and Ethics of Indian Constitution*, pp. 119–137. New Delhi: Oxford University Press.⏎
Ravish, Sushruth, and Sirola, Vikram Singh. 2023. 'Can Social Reflective Equilibrium Delineate Cornell Realist Epistemology?' *Philosophia*, 51: 2015–2033. https://doi.org/10.1007/s11406-023-00654-9⏎

Rawls, John. 1951. ‘Outline of a Decision Procedure for Ethics.’ *The Philosophical Review*, 60 (2): 177. https://doi.org/10.2307/2181696.⏎

Rawls, John. 1974. ‘The Independence of Moral Theory’, *Proceedings and Addresses of the American Philosophical Association*, 48: 5–22.⏎

Rawls, John. 1999. *A Theory of Justice* (Rev). Cambridge, MA: Harvard University Press.⏎

Rehren, Paul, and Blunden, Charlie. 2024. ‘Let’s Not Get Ahead of Ourselves: We Have No Idea If Moral Reasoning Causes Moral Progress’, *Philosophical Explorations*, 27 (3): 351–369. https://doi.org/10.1080/13869795.2024.2363876.⏎

Sauer, Hanno. 2019. ‘Butchering Benevolence Moral Progress beyond the Expanding Circle’, *Ethical Theory and Moral Practice* 22: 153–167. https://doi.org/10.1007/s10677-019-09983-9.⏎

Sauer, Hanno, Blunden, Charlie, Eriksen, Cecilie, and Rehren, Paul. 2021. ‘Moral Progress: Recent Developments’, *Philosophy Compass*, 16 (10): e12769. https://doi.org/10.1111/phc3.12769.⏎

Scanlon, Thomas. 2002. ‘Rawls on Justification’, in Samuel Freeman (ed) *The Cambridge Companion to Rawls*, pp. 139–167. Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521651670.004.⏎

Scanlon, Thomas. 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.

Tam, Agnes. 2020. ‘Why Moral Reasoning Is Insufficient for Moral Progress’, *Journal of Political Philosophy*, 28 (1): 73–96. https://doi.org/10.1111/jopp.12187.⏎

OceanofPDF.com

11 On the Content and Role of We-intentions

Olle Blomberg

DOI: 10.4324/9781003612827-12

1 Introduction

Joint intentional activities are ubiquitous in our lives. We cook, raise children, play games, fight for political change, engage in scientific and philosophical inquiry, plan and carry out bank heists, and intentionally cooperate for countless other purposes. Most language use is joint intentional activity (Clark 1996). Moreover, we typically acquire the skill to carry out an activity on our own by first doing it together with others. Besides being prevalent and instrumentally or intrinsically valuable, joint intentional activities are also normatively significant. They typically give rise to obligations and entitlements between participants and affect participants' backward-looking responsibility in significant ways (Blomberg and Hindriks 2020).[1] To understand human agency and sociality, we must understand joint intentional activity.

When we act together, we do not merely act strategically or in parallel. Contrast a group of friends snowboarding together with a group of strangers each snowboarding on their own in each other's proximity. Since everyone is trying to maintain an appropriate distance from the others, there will be coordination of turns and trajectories down the slope among friends and strangers alike. Indeed, in this pair of contrasting cases, it is not implausible that the behaviours of the agents involved would be indistinguishable, at least during a limited length of time. Since their behaviours may be indistinguishable, it is reasonable to think that some internal or mental

components must distinguish participation in joint intentional activity and acting in parallel with others (Searle 1990: 402–3).

Assuming a causal theory of action, I consider what the internal or mental component that distinguishes participation in joint intentional action could be.[2] According to the *content approach* I favour, joint intentional activity is distinguished by the fact that each participant's contribution is rationalised and guided by, roughly, an ordinary intention with the content 'that we *J*', where *J* is a joint activity (Bratman 2014). This yields a reductive account of a participant's *we-intention*.[3] After presenting my preferred such content account, I examine whether the content approach is able to explain how we-intentions can arise from so-called 'team reasoning' (Bacharach 2006). A team reasoning individual conceives of herself as a member of a group, deliberates about 'What should we do?', selects the best joint action *J*, determines her role within that action, and intends to perform that role.

Plausibly, joint intentional activities sometimes result from agents reasoning as group members in this way. Moreover, situations in which agents share a collective moral obligation often call for team reasoning (Schwenkenbecher 2021; Blomberg and Petersson 2024). Hence, it would be a significant limitation of the content approach if it failed to explain how we-intentions could be outputs of team reasoning. I argue that the content approach can explain this. When a team reasoner selects the best joint action *J*, she can form the intention that they *J* together. To the best of my knowledge, I am the first to explicitly argue for this possibility—a possibility denied by Björn Petersson (2017).[4]

In the next section, I explain why the content approach is promising. In Section 3, I present and motivate my 'Bratman-lite' account of we-intention, which builds on Michael Bratman's (2014) intention condition and my own earlier work (Blomberg 2016a; 2016b). In Section 4, I introduce team reasoning. In Section 5, I consider a potential argument for what I call the *Incompatibility Thesis*—the thesis that we-intentions on the content approach could not be outputs of team reasoning. According to the argument, gleaned in Abe Roth's (2014; 2020) work, such we-intentions do not belong to the same (sub-)type of mental state as ordinary intentions. In Section 6, I turn to Petersson's (2017) argument for the Incompatibility

Thesis, according to which we-intentions that arise from team reasoning have a we-perspectival feature that eludes the content approach. I argue that both arguments are unsound. In Section 7, I conclude that the content approach is compelling, even more so because it is compatible with team reasoning theory.

2 The Content Approach and Intending 'That We *J*'

Roughly, the friends' snowboarding is a joint intentional activity on the content approach because it is rationalised and guided by their ordinary individual intentions that they snowboard together. Each must intend that they *J* together. Other phrases can be used: 'intend to *J* with the other(s)' or 'intend to do their part of the *J*-ing', for example. But these phrases must then be understood such that the activity of the others is part of what is intended by each (Bratman 2014: 43–44).

Many think that intentions are necessarily concerned only with an agent's own actions. Given this 'own-action condition' (Bratman 1997), each friend could merely intend to do their part or perform their role, perhaps while believing the others each intend to do or perform theirs. But such intentions could not make the friends' snowboarding jointly intentional. After all, the snowboarding strangers who try to maintain an appropriate distance from one another could each intend to do what they realise is part of a larger pattern consisting of all their activities, and each could realise that each intends to do their part of this larger pattern (compare how each driver in a traffic jam might think of their own and the others' roles in it). In this case, there is no end that the strangers share, such that they each intend it. Such a shared (single) end is required for the agents' actions to compose a joint intentional activity.

What if each intends to merely do their part and desires that the others do theirs? This is also insufficient. I can desire that you do your part while nevertheless intending that you do not do it. What if we add a condition ruling out such obstructive intentions? Arguably, this would not yet capture typical everyday cases of joint intentional activity (cf. Asarnow 2020). Two strangers, who each intend to snowboard down the slope, desire that the

other do so in their vicinity and do not intend to obstruct the other, would not thereby be disposed to support or help the other when they fall or crash, say. But participants in everyday cases of joint intentional activity are normally disposed to help. This is because in such cases, each is committed to the joint activity rather than to merely do their part (Roth 2014: 280–3).

A possibility is to posit that the participants' we-intentions are a *sui generis* type of mental state, irreducible to ordinary intentions and other states. John Searle (2010: ch. 3) argues for this type of view because he accepts something like the own-action condition on the content of intentions (2010: 53, 56).[5] According to Searle, a we-intention is an intention distinct in kind from an ordinary individual intention, in that it is not subject to the own-action condition. I call this the *mode approach*.[6]

However, the own-action condition should be rejected (Bratman 1997; 2014: 60–64; Núñez 2022; for the contrary view, see e.g. Kietzmann 2018 and Schmid 2023: 33–38).[7] Hence there is no need to take the mode approach. (In Section 5, I address another potential motivation for taking the mode approach.) An agent can rationally intend that they—she and the others—carry out a joint activity *J*. In some cases, she can do this because she can make reasonable assumptions or conditional predictions about what the others will intend and do if she manifests to them her intention that she and they *J* together (Bratman 1997; 2014: 73–75; Ludwig, 2016: 208–10). These assumptions or predictions enable rationally intending that they *J* together in the same way assumptions and conditional predictions about non-agent parts of one's environment can enable rational intending of one's own activity. I can rationally form an intention to grow tomatoes partly because I can reasonably assume my local garden centre will have the appropriate seeds and planting soil for me to acquire when the time is right. This intention to grow tomatoes is also partly enabled by my conditional prediction that if I were to form and act on such an intention, the seeds I would put in the soil would grow into tomato plants. Thanks to such assumptions and conditional predictions, I can settle that the seeds will grow into tomato plants, and hence rationally intend to grow them. Similarly, I may be able to settle that we will dance together, thanks to my conditional prediction that you will intend that we dance together if I manifest to you my intention that we dance together. I can then rationally

intend that we dance together. There is nothing mysterious about this. As Bratman (1997: 58) puts it: ‘This is not an unusual form of control of your action, but just the ordinary predictability of ordinary agents’.

Freed from the own-action condition, each participant’s ordinary intention that they *J* together can rationalise and cause their *J*-ing. For these intentions to provide the foundation of a reductive content account of we-intentions, the participants cannot be required to conceive of the intended *J*-ing as a joint intentional activity. This would make the account problematically circular (Petersson 2007). However, a joint activity need not be a joint *intentional* activity. Indeed, the parallel activity of the strangers snowboarding is a kind of joint activity, albeit not a joint intentional activity. Hence, when it comes to snowboarding together, there is a conception of several agents snowboarding that covers both the friends’ and the strangers’ snowboarding activity. This conception can figure in the intention of each of the snowboarding friends (Bratman 2014: 44–48).[8]

Due to its parsimony and explanatory power and consistency with widely accepted causal theories of action, a content approach based on ordinary intentions that they *J* together should be the default approach for making sense of joint intentional activity. In the next section, I present my own preferred version of such an account.

3 A Bratman-Lite Account of We-Intention

For each to have the relevant we-intention, it is not sufficient that each intends that they *J* together. This is partly because joint intentional activity is intentionally cooperative (Ludwig 2016: 225–8; Blomberg and Malmqvist 2024). Intending that oneself and others *J* together is compatible with viewing and treating them as mere social tools. Suppose I threaten you with a gun to ensure that you do your part of our *J*-ing, while you threaten me in the same way (Pacherie 2013: 1822; Bratman 2014: 49; Butterfill 2015: 152). My intention that we *J* could in this case be satisfied if your behaviour combines with my intentional activity such that we *J*—you need not do your part effectively guided by your intention that we *J*. And vice versa for your intention with respect to my activity. If our *J*-ing were the

result of such intentions, it would not be intentionally cooperative (Ludwig 2016: 195–7).

We can ensure that a we-intention is cooperative by endorsing Bratman's *intention condition* (for alternative conditions ensuring this to different extents, see Ludwig 2016: ch. 13; Butterfill 2015, 2016; and Blomberg 2021). According to this condition, if you and I have a 'shared intention', then 'we each have intentions that we *J*; and we each intend that we *J* by way of each of our intentions that we *J*, … and … by way of sub-plans that mesh' (Bratman 2014: 152). This condition rules out the kind of coercion case just described, as well as some forms of deception. Suppose I intend that we snowboard down to the bottom of the cable car by way of each of our intentions that we do so, and by way of co-realisable sub-plans. I cannot then also rationally intend to trick you into taking the black (difficult) slope down if I know that your sub-plan is for us to take the green (easy) slope down. If I were to trick you down the black slope, I would frustrate my own intention that we ride down by way of co-realisable sub-plans. Hence, the core of a we-intention is plausibly an ordinary intention that we *J* by way of each of our intentions that we *J*, and by way of meshing sub-plans.

Bratman does not take his intention condition to ensure that each agent is '*cooperating with* the other' in the resulting joint intentional activity (2014: 38). This is because the condition can be satisfied even if I coerce you to form or acquire the required intention. In contrast to Bratman, I think that, even in such a coercion case, if our intentions result in joint intentional activity, then we are intentionally cooperating—albeit within the confines of a coercive interaction (Blomberg and Malmqvist 2024). This is analogous to how competition is embedded in cooperation when we play chess, for example.

Despite this, having the complex intention required by Bratman's intention condition does not quite suffice for having a we-intention. Notice that the various participants may each represent their shared end differently (Blomberg 2016a):

> *Forest Hunters*: You and I have teamed up for a forest hunt, relying on our complementary abilities—my sharp hearing compensates for your deafness, and your keen vision compensates for my blindness. We are each aware that a single animal lurks in the canopy above. It rustles

the leaves and casts a shadow on the forest floor. I intend that we catch the animal rustling the leaves ('Rustle' refers to the animal when conceived in this way). You, on the other hand, intend that we catch the animal casting the shadow ('Shadow'). I intend that we catch Rustle by way of each of these intentions and by way of sub-plans that mesh. You intend that we catch Shadow by way of each of these intentions and by way of sub-plans that mesh.

Here the intention condition is satisfied, and we have a shared end: that we catch Rustle/Shadow. You and I each also believe that we have a shared end. We represent this shared end differently, but this does not preclude us from each having, and acting on, a we-intention. Now, Bratman makes 'the simplifying assumption that […] the participants will have a common conception of *J*, that there is in this way a match in what is intended by each' (2014: 42). But: 'Perhaps we can share an intention to go to NYC if I intend that we go to the city that is the home of the Yankees and you intend that we go to the city that is the home of the Mets' (2014: 42). Forest Hunters illustrates that the participants' intentions indeed need not represent their shared end in the same way, and that participants need not believe or assume that their intentions represent the shared end in the same way. Given our awareness of our respective disabilities in Forest Hunters, I do not believe *de dicto* that we each intend that we catch the animal rustling the leaves, nor do you believe *de dicto* that we each intend that we catch the animal casting the shadow.

Now the reason for adding something to Bratman's intention condition comes into view. Consider the following case:

Frege-Style Hunters: You and I are out hunting in the forest. We both possess keen hearing and vision. In the nearby canopy, a single animal rustles the leaves and casts a shadow on the forest floor. However, we mistakenly believe that the rustling creature ('Rustle') and the shadow-casting creature ('Shadow') are distinct, falsely assuming that there are two animals. I want us to catch Rustle, while you want us to catch Shadow. I believe that we can only capture Rustle if it is distracted by your pursuit of Shadow, and you hold a reciprocal belief about catching Shadow. We are aware that each of us

> has this desire and belief. Given this, I anticipate that if you were to recognise that I intended that we catch Rustle, then you would intend that we catch Shadow. Likewise, you anticipate that if I were to recognise that you intended that we catch Shadow, then I would intend that we catch Rustle. Accordingly, I form the intention that we catch Rustle by way of each of these intentions and by way of sub-plans that mesh. You likewise form the intention that we catch Shadow by way of each of these intentions and by way of sub-plans that mesh.

As in Forest Hunters, the intention condition is satisfied, and we have a shared end: that we catch Rustle/Shadow. But in Frege-Style Hunters, we each mistakenly believe that this end is not shared. Much like an agent in a so-called 'Frege case' fails to realise that certain expressions co-refer (Fodor and Lepore 1992: 167), we fail to realise that our intentions are directed at the same token end. (Nevertheless, our intentions may have been rationally formed. If it seems otherwise, recall that 'that we catch Rustle/Shadow' need not be equivalent to 'that we jointly intentionally catch Rustle/Shadow'. To avoid circularity, it might be equivalent to something like 'that Rustle/Shadow is caught as the combined result of our intentional actions'.) Each of us views the other's intention as a means by which we can pursue our own end, which we wrongly take to be distinct from the other's. Neither of us therefore has a we-intention to catch Rustle/Shadow. Hence, having the complex intention specified by Bratman's intention condition is not sufficient for having a we-intention.

What needs to be added is simply an assumption or belief that the end that we *J* is shared, such that we each intend it (Blomberg 2016a; see also Searle 2010: 52–53). With this *doxastic shared end condition* in place, Bratman's simplifying assumption that participants have a common conception of *J*—i.e., beyond their conception of *J* as the end they share, such that they each intend it—becomes superfluous.[9] The condition accommodates Forest Hunters and excludes Frege-Style Hunters as examples of joint intentional activity.

By contrast, a condition requiring belief (or assumption, knowledge, or common knowledge) that each of us intends that we *J* (by way of … and so on) cannot accommodate Forest Hunters and exclude Frege-Style Hunters.

The belief ascription must be understood *de re* to accommodate Forest Hunters, so that believing that I intend that we catch Rustle and that you intend that we catch Shadow counts as believing that we intend the same token end. But it must be understood *de dicto* to exclude Frege-Style Hunters, so that believing that I intend that we catch Rustle and that you intend that we catch Shadow counts as believing that we intend distinct ends. In Frege-Style Hunters, I do believe *de re* that we each intend that we catch Rustle, and you believe *de re* that we each intend that we catch Shadow. To avoid this dilemma, the condition must require *de dicto* that I believe that we each intend *de re* that we catch Rustle, and require *de dicto* that I believe that we each intend *de re* that we catch Shadow. Each of us satisfies this in Forest Hunters, but neither satisfies it in Frege-Style Hunters. Fortunately, the doxastic shared end condition makes such a convoluted condition unnecessary.

Is anything else needed? No. Many have assumed participants must have common knowledge of each other's intentions. But while such common knowledge may be necessary for rational we-intending in some circumstances, as well as for enabling shared deliberation and shared planning about how to carry out the we-intentions (Bratman 2014: 57; cf. Roy and Schwenkenbecher 2021), such strong cognitive glue is not necessary for we-intending and joint intentional activity as such (Blomberg 2016b; Ludwig 2016: 196–7, 219–21).[10]

Perhaps attitudes can be combined in other ways to amount to a we-intention. But my claim is that Bratman's intention condition and my doxastic shared end condition each specify an insufficient part of jointly sufficient conditions on having a we-intention. We thus get this minimally sufficient pair of conditions for an agent, 'I', to have a we-intention to *J*:

1. I intend that we *J*, by way of each of our intentions that we *J*, and by way of sub-plans that mesh.
2. I assume or believe that the end that we *J* is shared, such that we each intend it (by virtue of the intentions ascribed or mentioned in condition 1).

These conditions ensure that the we-intention has a cooperative character, and that the end the agent intends to bring about is assumed to be shared. If

we each have a we-intention to *J* and these appropriately guide the resulting joint *J*-ing, then we will jointly intentionally *J*. I call this my 'Bratman-lite' account, since its crucial first condition is taken straight from Bratman's account.

Some readers might balk at 'lite'. After all, the account requires the ascription of an intention with complex self-reflexive content, plus an assumption or belief about that intention. One might think it is implausible to ascribe such states to each of, say, two people who are having a conversation or are trying to fit a large sofa into a narrow elevator. Or one might think it is implausible to ascribe such states to children acting together (Pacherie 2013). And some habitual or highly skilled joint behaviours may be basic joint actions, wherein each participant directly intends the joint bodily movement (Blomberg 2011): consider two people shaking hands, or two professional figure skaters' joint move in their dance routine on the ice. Ultimately, it is an empirical question what is involved in the production and guidance of any coordinated interaction. However, the intention and the assumption or belief required by conditions 1 and 2 can be merely implicit or tacit (Bratman 2014: 104). When they are, the agent need pay no heed to them, and they need play no role in the agent's conscious thinking and deliberation. In that case, the we-intention would only be virtually present—'in the co-pilot's [seat], ready to assume control' (Pettit 1995: 322). As long as the agent subconsciously monitors for any signals indicating that her ordinary routine behaviour is going off-track, and is disposed to react appropriately if this occurs, then her activity can be said to be guided, virtually at least, by the intention and the assumption or belief required by conditions 1 and 2.

4 We-Intentions, Practical Reasoning and Team Reasoning

The considerations I have appealed to in support of my Bratman-lite account are all concerned with the success conditions of a we-intention and how these rule out rationally combining a we-intention with certain coercive or deceptive intentions. This should come as no surprise. Intention's functional role tends to be characterised exclusively in terms of

inputs to, and constraints on, downstream or concurrent events and processes. For example, unlike simple combinations of desires and means-end beliefs, intentions settle practical deliberative questions, pose problems for further means-end reasoning and planning, and tend to ensure that the agent acts on time and when appropriately situated (Bratman 1987).

However, intentions must also be the output of various 'upstream' processes and mechanisms, such as practical reasoning, which take the agent from her pro- and con-reasons for available courses of action to her intention (1987: 58). Even when spontaneously acquired, intentions are typically outputs of some reasons-responsive mechanism. Consider a sailor who ducks when seeing the boom swinging towards her across the cockpit. If her ducking is controlled by an intention, then the intention was presumably not the output of reasoning or deliberation. But its acquisition may nevertheless be responsive to her reasons.

Intentions are also '*reason-preserving*' (1987: 63). This is clearest in the case of future-directed intentions: Unless the agent reconsiders, she will act for the motivating reasons that originally led her to intend the action, even if those reasons are no longer present to mind or have been forgotten (Roth 2020: 174). In this way, intentions are tied both to prior reasoning and to motivating reasons.

A we-intention, on a content account, can play these roles as outputs: My intention that we dance together can be the output of my practical reasoning regarding what I should bring us to do, where this reasoning is done in light of what I think would be most rewarding for us to do, and in light of my conditional prediction that you would intend that we dance if you were to recognise that I intended that we dance, and so on. And my intention that we carry your large and heavy suitcase off the train together may be spontaneously acquired when I see you struggling to lower it onto the platform by yourself. We-intentions can also be outputs of our 'shared deliberation', where this is itself a joint intentional activity where we discuss whether and how to proceed with some complex joint endeavour, such as a scientific or philosophical research project (Bratman 2014: ch. 7).

However, it has been argued that a we-intention, on the content approach, cannot be the output of 'team reasoning'. Unlike shared deliberation, team reasoning is a form of practical reasoning done by an individual agent (Sugden 2000; Bacharach 2006; Colman and Gold 2018).

Instead of being prompted by the first-person singular deliberative question 'What should I do?', where the agent seeks the best response to others' expected behaviour, team reasoning is prompted by the first-person plural deliberative question: 'What should we do?' Provided there is a unique answer, this first step of team reasoning yields a judgement regarding what the team, 'we', should do. In a second step, the agent then considers what her role in that joint action is, yielding a judgement about what she herself, as a team member, should do.

Theories of team reasoning tend to assume, with mainstream decision and game theory, that a rational agent can be modelled as having only expectations (or beliefs) and preferences (desires). But this model can in principle be augmented with intentions (see Hakli, Miller and Tuomela 2010). With such an augmentation in place, it is reasonable to take the outputs of both steps of team reasoning to be some kind of intention. In the first step, upon reaching the judgement that the team should *J*, one might think that the agent could form an intention that they *J* together (by way of … and so on); in the second step, she could form the intention to do her part of, or perform her role in, the joint action. However, if one regards intentions motivated by collective reasons as distinct in kind from those motivated by individual reasons, one will resist the idea that the outputs of team reasoning could be ordinary individual intentions that they *J* together, at least given the assumption that team reasoning engages only with collective reasons, while first-person singular practical reasoning engages only with individual reasons. Furthermore, Petersson (2017) argues that we-intentions produced by team reasoning possess a we-perspectival feature that is part of their mode and that therefore eludes the content approach. I examine these arguments for the Incompatibility Thesis in Sections 5 and 6.

One reason for thinking we engage in team reasoning is that it provides a simple and attractive explanation of how we find the obvious and only rational solution in what is known as the Hi-Lo game. To illustrate, consider the *Footballers' Problem* (Sugden 2000): Two football players are trying to pass the ball between them. In the heat of the game, they cannot communicate. The pass can be made to the left or to the right. A pass to the left would be preferable (the 'Hi' Nash equilibrium). A pass to the left will be brought about if Player 1 kicks the ball to the left while Player 2 runs to the left to receive it. A pass to the right would be the second-best outcome

(the ‘Lo’ Nash equilibrium). A pass to the right will be brought about if Player 1 kicks the ball to the right while Player 2 runs to the right to receive it. Least preferable is a failure of coordination. For a game-theoretical representation of this situation, with the numbers representing the players’ utilities, see Table 11.1.

Table 11.1 Payoff matrix for the Footballers’ Problem (Hi-Lo)⏎

	Player 2	
Player 1	Run to the left	Run to the right
Kick to the left	10, 10	0, 0
Kick to the right	0, 0	5, 5

Orthodox game theory provides no determinate recommendation for the players here. Each is supposed to choose the best reply to whatever she believes the other will do. If the other plays left, then play left; if the other plays right, then play right. But whether either player plays left or right depends on what she thinks the other player will do. No factor can rationally tip either player’s expectations about whether the other player will play left or right. Could they not each assume the other is as likely to play left as right, and then play left because it uniquely maximises expected utility? Assuming this would be to treat the other as irrational, and this should not be required to solve the problem. After all, for each player, kicking or running to the left is intuitively the obvious rational choice, and they each know this (Colman and Gold 2018: 1773–4).

The problem is not that the players are acting egoistically. Altruistic ‘team benefactors’—best-reply reasoners who each ask ‘What should I do for us?’—cannot rationally solve the game (Petersson 2017: 202–3). It is natural to think each player on a football team evaluates the outcomes in relation to her ‘team preferences’, that is, to her team’s standard of success (Sugden 2000). A player’s team preferences need not be aligned with her individual preferences. For example, what works best for the team’s goal of winning the match, or remaining in the league, need not be what works best for a football player’s personal career aspirations. Similarly, if I am one of the friends snowboarding together, I may reason and act based on what the

group wants, such as when I try to figure out a route down to base camp that takes into account everyone's skill level, time constraints and scenic preferences, although I may personally want us to take the shortest and steepest route. Team reasoning theorists assume that players share a single team utility function, typically illustrated by a function that equates team utility with the average of the players' individual utilities. However, that the players have such a shared team utility function does not itself avoid the deadlock of Hi-Lo. (Table 11.1 accurately represents the team benefactors' situation.) As long as each player merely seeks the best response to the others' expected behaviour or choice, they will remain stuck.

According to team reasoning theory, what is needed for the two footballers to overcome their deadlock is a 'transformation of reasoning' (Bacharach 2006: 90). The unit of agency presupposed in each player's practical reasoning must change from herself, considered individually ('I'), to the two of them together, considered as a team ('we'). Table 11.2 represents the Footballers' Problem after such a reasoning transformation, with the numbers representing the shared team utility function.

Table 11.2 Payoff matrix for the Footballers' Problem (Hi-Lo), as faced by team reasoners⏎

	Player 2	
Player 1	Run to the left	Run to the right
Kick to the left	10	0
Kick to the right	0	5

In team reasoning's first step, the player selects the outcome that is best for the team (the upper left corner in Table 11.2's payoff matrix). If the player is an agent with intentions in addition to her preferences and expectations, then this selection can, I claim, result in an intention that the players together bring about this outcome ('that we pass the ball to the left'). In the second step, the player then concludes that she should do her part of the profile of actions most likely to bring about that outcome (kick or run to the left), arriving at an intention to do this part. In the Footballers'

Problem (Hi-Lo), this second step is swift and obvious, but it need not be. Suppose two colleagues are tasked with organising an international research workshop, and they can coordinate either on holding it on-site at their home department (Hi), or on having it hosted online by a web conferencing platform (Lo). In this case, further practical reasoning, planning, bargaining, and negotiation may be required for each to transition from an intention that they bring about Hi to an intention to do their own part of the action profile most likely to bring about Hi (Pacherie 2013: 1836). Moreover, an intention of each that they jointly bring about the selected outcome may raise the question 'What should we do to bring it about?', prompting further team reasoning.

Team reasoning requires the rejection of what Christopher Woodard (2011: 263) calls the mirroring assumption about reasons for action.[11] For our purposes, we can understand it as pertaining directly to practical deliberation in the following way:

The mirroring assumption: The agent or agents who deliberate always match—and determine—the agent or agents of the actions being considered.

If this assumption were true, then only *I* could ever practically deliberate about *my* actions, and only *we* could ever practically deliberate about *our* joint actions. In team reasoning, each agent separately deliberates about what they should do together, responding to reasons for or against action alternatives that involve contributions from herself and the others. Thus team reasoning violates the mirroring assumption. (In contrast, shared deliberation is in accordance with it. Shared deliberation might unfold as follows: I raise the question 'What should we do?' aloud, out in the open between us. There is uptake of this question on your part, and we start to deliberate about what we ought to do. This deliberation is realised through a conversation between us. In this process, the posing of the question, the establishment of the audience's uptake, and the deliberation that follows are all part of a joint intentional activity carried out by *us*. Hence, in line with the mirroring assumption, the agents who deliberate, 'we', are identical to the agents of the intended action our deliberation results in, i.e. 'us'.)

How might team reasoning in each agent get started? Michael Bacharach (2006) proposes that team reasoning is prompted by *group identification*.

As a result of such identification, an agent frames the decision problem as a problem for 'us', herself and the others considered collectively. The notion of group identification is taken from the social identity approach in social psychology (see e.g. Hogg, Abrams and Brewer 2017). According to Bacharach, an agent does not voluntarily choose whether to identify with a group. Rather, group identification is arationally triggered by circumstances and situational cues that make group identity salient (cf. Sugden 2000). This does not imply that team reasoning is not rational. Since tendencies to group identify are likely to be widely shared in certain types of situations, a team reasoner will often be warranted in presupposing that others are team reasoners too. Bacharach speculates, for example, that the strong interdependence that exists between the individuals' interests in social dilemmas such as Hi-Lo is a universal and reliable group-identification cue (2006: 143–4).[12]

I simply take Bacharach's team reasoning theory and the rejection of the mirroring assumption for granted here (for critical discussion, see e.g. Hindriks 2012; Roth 2014; Browne 2018). I find the proposal that we often group identify, and often tacitly or explicitly team reason, plausible. Since situations calling for joint action will often resemble Hi-Lo in having a scope of mutual advantage for coordinating agents, joint action will often involve group identification and team reasoning. Furthermore, when agents face a *moral* Hi-Lo problem they often have a collective moral obligation to bring about the moral Hi equilibrium (Schwenkenbecher 2021; Blomberg and Petersson 2024). It is therefore important for an account of we-intentions to be compatible with team reasoning theory. In the next two sections, I consider and reject two arguments according to which the appropriate outputs of team reasoning cannot be we-intentions given a content approach.

My conclusion will be that there is no reason to think that my Bratman-lite account of we-intention is incompatible with team reasoning theory. As a result of the first step of team reasoning, an individual can rationally form an intention that she and the others *J*, by way of each of their intentions that they *J* together, and by way of co-realisable sub-plans. My claim is not that the first step of team reasoning invariably results in such an intention, but that it can result in it. If an agent does form such an intention, then the

presumption that the others are team reasoners with the same team utility function will make it reasonable for her to assume that the end that they *J* together is shared, such that she and the others each intend it. She will then have a we-intention and be under rational pressure to intend to perform her own role within the *J*-ing in a way that is sensitive to the others' intentions and sub-plans (Bratman 2014: 64; Ludwig 2016: 103–4). A transition from the we-intention to the intention to perform her role corresponds to the second step of team reasoning.

5 Collective Reasons, the Mode Approach, and the Incompatibility Thesis

It might be tempting to think that team *reasoning* and team *reasons* are tightly connected. In the Footballers' Problem (Hi-Lo), each team reasoning player can act on the consideration that the best she and the other player can do to further their team's aim of winning is to pass the ball to the left. If each instead engages in best-reply reasoning, then this is not a conclusive consideration in the light of which either of them can rationally act. This may suggest such collective reasons are only available to agents who group identify and team reason.

Combined with another premise, we can get an argument for the Incompatibility Thesis. Consider Roth's (2020: 175 n. 8) claim that 'intentions formed when one undertakes team reasoning are quite distinct from ordinary individual intentions'; they are not simply ordinary individual intentions 'with a collective subject matter'. This, he tells us, is because they preserve collective motivating reasons rather than 'proprietary reasons', that is, individual motivating reasons (2020: 174–5). Roth's focus here is the intentions that are the upshot of the second step of team reasoning—intentions to do one's part—but his point would also apply to intentions that are the upshot of team reasoning's first step—intentions 'that we *J*'. It is not entirely clear what Roth means when he writes that intentions formed by team reasoning are 'quite distinct'. But one possible interpretation is that he takes them to belong to a distinct (sub-)type of mental state from ordinary individual intentions—that they are we-mode intentions—because they are the upshot of distinct forms of reasoning or

reasons-responsive processes (for support for this interpretation, see Roth 2014: 286 n. 19). In line with the Incompatibility Thesis, this would imply that the outputs of the first step of team reasoning cannot be ordinary individual intentions 'that we *J*'. While my proposed content account might accurately spell out the satisfaction conditions of we-intentions formed via team reasoning, these we-intentions would belong to a different type of intention than ordinary individual intentions.

This argument would fail for two reasons. First, intentions arising from best-reply reasoning can preserve collective motivating reasons, as the possibility of best-reply reasoners who act on team preferences—'team benefactors'—demonstrates. To illustrate, consider the Footballers' Problem again: Suppose Player 1 knows (for whatever reason) that Player 2 intends that they together perform a pass to the left. Player 1 can then reliably predict that Player 2 will perform her role in this joint action, i.e. run to the left. Furthermore, suppose Player 1 is a team benefactor: her deliberation is prompted by the question 'What should I do for us?' rather than 'What should we do?'. Player 1 can then rationally intend that she and Player 2 together perform a pass to the left. Despite being the output of best-reply reasoning, this intention would preserve the collective motivating reason for them to perform a pass to the left—i.e. the consideration that it maximises the team's chances of scoring a goal. Such an intention 'that we *J*' would be an 'ordinary individual intention', in the sense that it was formed by best-reply reasoning but would belong to the same (sub-)type of mental state as a team reasoner's intention 'that we *J*'. Perhaps this possibility is what Roth has in mind when he writes that '[t]here may be some overlap' between the reasons available in team reasoning and in best-reply reasoning (2020: 175). At any rate, this shows that even if preserving collective motivating reasons were to make a class of intentions distinct, the Incompatibility Thesis would not follow.

Secondly, it would be a mistake to take the preservation of collective motivating reasons to mark a distinct type of intention. Intentions are relatively fixed points in our mental economy: they screen off prior deliberation and postpone further deliberation about filling in partial plans until later. In tying we-intentions and ordinary individual intentions to collective and proprietary reasons, we would make it impossible to describe participants as having the same intention while preserving their distinct

kinds of motivating reasons. In the Footballers' Problem, suppose Player 2 group identifies with her and Player 1's team, and then team reasons and intends that the two of them together perform a pass to the left. Her motivating reason is that this is the best option for their team. Suppose Player 1 knows all this. Player 1 can then, through best-reply reasoning, form a matching intention that they together perform a pass to the left. When she does this, suppose she only considers what is best for her. She does not care whether her team wins; her focus is on her own reputation and the audience's adoration of her alone. Here we should be able to say that Player 1 and Player 2 have matching we-intentions to perform a pass to the left, even though their motivating reasons differ in kind. By tying we-intentions and ordinary individual intentions to collective and individual reasons, we would also make it impossible to describe someone as having retained her intention while changing her motivating reason from one kind to another. In the case just described, suppose Player 1 suddenly becomes motivated by the consideration that passing the ball to the left would be best for her team—she switches from being egoistic to being 'groupish', as it were. Here, her intention that they perform a pass to the left would be retained even though the kind of motivating reason she has for enacting the intended joint action has changed. To allow for this, intentions preserving distinct kinds of motivating reasons must belong to the same type of intention.

We should reject both premises of this argument for the Incompatibility Thesis: ordinary individual intentions can preserve collective motivating reasons, and intentions that preserve distinct kinds of motivating reasons should not count as distinct types of intentions. For all we have seen, ordinary intentions 'that we *J*' can be outputs of team reasoning. Next, I turn to Petersson's (2017) explicit argument for the Incompatibility Thesis, according to which ordinary intentions 'that we *J*' cannot be such outputs.

6 Petersson's Argument for the Incompatibility Thesis

Petersson's compressed argument for the Incompatibility Thesis is embedded in his following critical discussion of Bratman's account of

shared intention:[13]

> Unlike ‘we intend to *J*’ phrases of the form ‘I intend that we *J*’ are not commonly used in ordinary language, but if there is a question to which such a phrase is the answer, this seems to be a question I may ask myself when I am about to form my intention concerning us. Like in the intention that is expected to result from this deliberative process, ‘we’ figures in the *content* of the question but it is asked from my *perspective* rather than ours. In this framework, my intention that we do this rather than that would presumably result from what I want for us, my caring for how well we do. So, it seems reasonable to assume that the Bratmanian co-operator would not be asking what *we* should do in Bacharach’s distinct sense. Being a Bratmanian co-operator would not help in the potential Hi-Lo game.
>
> […T]he explanation of [the Bratmanian co-operator’s] failure [at team reasoning] is that the group merely figures in the content of her attitudes, and that this content is conceived from an individual perspective.
>
> (Petersson 2017: 205)

Petersson makes two distinct claims. One claim is that ‘[b]eing a Bratmanian co-operator would not help in the potential Hi-Lo game’. Regarding this claim, he clarifies at the end of the page that ‘Bratman’s shared intention will not suffice to avoid the deadlock’ that best-reply reasoners, including team benefactors, face in Hi-Lo (2017: 205–6). Here the Bratmanian co-operator’s intention ‘that we *J*’ is supposed to be upstream of her practical reasoning in Hi-Lo. To illustrate, suppose Player 1 and Player 2 in the Footballers’ Problem each intend that they pass the ball between them. They must now figure out whether they should pass it to the

left or to the right. In this sort of situation, best-reply reasoning within the framework of a Bratmanian (or Bratmanian-lite) shared intention would not get the parties out of the Hi-Lo deadlock (see also Hakli and Mäkelä 2017: 133–6; for the contrary view, see Ludwig 2017). I agree.

The second, more tentative claim is that a Bratmanian we-intention cannot be the output of team reasoning. This is the Incompatibility Thesis. Petersson observes that if an intention 'that we *J*' is the upshot of deliberation, then what prompts this deliberation 'seems to be a question I may ask myself when I am about to form my intention concerning us' (2017: 205). He suggests that the only such deliberative pro-group question the Bratmanian co-operator may ask herself is the team benefactor's question: 'What should I do for us?' In this question, '"we" figures in the *content* of the question but it is asked from my *perspective* rather than ours' (2017: 205). However, team reasoning is usually informally presented as a form of practical reasoning prompted by a question that also has 'we' in its content: 'What should *we* do?' Since this question is asked by an individual, it is in some sense also asked from 'my' perspective. What, then, is the Bratmanian co-operator's problem?

Understanding the supposed problem requires some familiarity with Petersson's 'perspectival' account of we-intentions. Petersson argues that, to capture the nature of we-intentions, we need to distinguish between *the subject of intention* and *the ontological subject* who has the intention (2017: 212; see also Mathiesen 2002). The subject of intention is the implicit agent perspective—'I' or 'we'—from which the intention is held. It is a perspectival feature of the mode of intending rather than part of *what* is intended (see also Schmitz 2017). While it participates in determining the intention's success conditions, it is not part of the intention's content.[14] When an agent has a we-intention, that agent is the ontological subject, but the subject of the intention is 'we'.

Petersson seems to suggest that just as we should distinguish the subject of intention from the ontological subject, we should distinguish the subject of reasoning from the ontological subject who performs the reasoning. The difficulty for the Bratmanian co-operator is supposed to be that 'we' at most figures in the content of her deliberative question. According to Petersson, a transformation from best-reply reasoning to team reasoning requires that the

subject of reasoning changes from 'I' to 'we' (while the ontological subject remains the individual agent).

Petersson's claim, then, is that 'we' can never be the subject of the Bratmanian co-operator's practical reasoning. Unlike the team reasoner, the Bratmanian co-operator cannot ask from the we-perspective what she and the others should do. In a sense, this is trivially true. Bratman and others who do not distinguish between the subject of reasoning/intention and the ontological subject cannot claim that the subject of reasoning is 'we' while the ontological subject is the individual agent. However, for it to work directly against their accounts, Petersson's argument must not depend on his distinction between the subject of reasoning and the ontological subject.

Since Bratman does not assume that participants can team reason, and only offers an account of shared deliberation, it might be tempting to think that he should accept the mirroring assumption. Recall that according to the mirroring assumption, only *I* can practically deliberate about *my* actions, and only *we* can practically deliberate about *our* joint actions. If Bratman accepted this assumption, then a Bratmanian co-operator could not herself form intentions 'that we *J*' via practical deliberation about what 'we' should do. But it seems unlikely that a champion of the idea that 'I' can intend 'that *we J*' would insist that 'I' cannot practically deliberate about 'our' activity. Indeed, given the intimate link between intentions and practical deliberation, rejecting the own-action condition and rejecting the mirroring assumption ought to go hand in hand. If anything, the content approach and team reasoning theory should be natural allies.[15]

However, one might think that rationally forming intentions 'that we *J*' and engaging in team reasoning require distinct and incompatible stances toward one's co-participants. Recall a claim from Section 2: An agent can rationally intend 'that we *J*' if she can make reasonable assumptions or conditional predictions that the others will do their parts of the *J*-ing if she manifests her intention 'that we *J*' to them (Bratman 1997; 2014: 73–75; Ludwig, 2016: 208–10). This involves a predictive stance toward the others' activity that is different from, and seemingly incompatible with, the stance presupposed by team reasoning. When engaged in team reasoning, an agent deliberates directly about what 'we' should do. She takes a deliberative stance toward not only her own but also the others' actions.

Given this apparent tension between stances, one might think that intentions 'that we *J*' cannot arise from team reasoning. However, Bratman and Kirk Ludwig are giving accounts of we-intentions and joint intentional activity made for agents who lack, or at least do not rely on, capacities for team reasoning. They thus offer arguments for how best-reply reasoners can rationally intend 'that we *J*'. What we are now considering is whether intentions 'that we *J*' can be outputs of team reasoning, given that this form of practical reasoning is on the table. On the face of it, team reasoning would thus provide a new route for agents to rationally form intentions 'that we *J*'. An alleged incompatibility between predictive and deliberative stances cannot be used to support the Incompatibility Thesis.

Petersson's argument could be understood, more modestly, as premised on his perspectival account of intention and reasoning. There might be reasons for accepting a perspectival account that have nothing to do with team reasoning, and once it is accepted, the Incompatibility Thesis might follow. Petersson's claim would then be that an *I-perspectival* intention 'that we *J*' could not arise from team reasoning; only a we-perspectival intention could. If the Bratmanian co-operator's intention 'that we *J*' is I-perspectival—such that the 'we' is *only* part of what is intended—then it could not result from team reasoning. This argument would seem to assume a perspectival analogue of the mirroring assumption:

The perspective mirroring assumption: In any practical reasoning that leads to an intention, the subject of reasoning always matches—and determines—the subject of intention.

Deliberative intentions resulting from team reasoning would always be we-perspectival, whereas those resulting from best-reply reasoning would always be I-perspectival.

I have two critical questions about Petersson's argument thus interpreted. First, why should we assume that the Bratmanian co-operator's intention 'that we *J*' should be understood as I-perspectival? After all, if the intention is the output of deliberation prompted by the question 'What should we do?' in a situation framed as a decision problem for 'us', then it is appropriate to say, in at least a loose sense, that the intention is held from a we-perspective. The fact that Bratman does not construe the mode of intending as having an agent-perspectival feature does not imply that intentions 'that we *J*' should be understood as I-perspectival rather than we-

perspectival if such a feature were added. Secondly, why should the perspective mirroring assumption be accepted? If team reasoning is a valid form of practical reasoning, then the mirroring assumption is false. And if the mirroring assumption is false, then why should we not think the perspective mirroring assumption is false with it? Insofar as the perspective mirroring assumption has any plausibility, it seems to piggyback on the plausibility of the mirroring assumption.

In the absence of any interpretation of Petersson's argument that makes it compelling, we should reject the Incompatibility Thesis regarding the content approach to we-intentions and team reasoning theory. Those who take the content approach can thus appeal to team reasoning theory to explain how people rationally form we-intentions in some situations.

7 Conclusion

According to the content approach, a joint activity *J* is a joint intentional activity if and only if it is appropriately caused and coordinated by the participants' we-intentions to *J*, where these consist of their ordinary mental states with collective contents. Drawing on Bratman's intention condition, I have argued the following two pieces suffice for making up a participant's we-intention to *J*: First, an intention that she and the others *J*, by way of each of their intentions that they *J* together, and by way of co-realisable sub-plans. Secondly, an assumption or belief that the end that they *J* together is shared, such that they each intend it. This is my Bratman-lite account of we-intention.

I have emphasised that an account of we-intentions must explain not only how they rationalise and appropriately cause downstream joint activity, but also how they can be outputs of different forms of practical reasoning. One form of practical reasoning that leads people to join forces and act together is team reasoning. Those who have given reductive content accounts of we-intentions have tended not to embrace team reasoning as a valid form of practical reasoning, though. Moreover, Petersson has argued that we-intentions based on ordinary intentions 'that we *J*' could not be outputs of team reasoning. I have argued that this Incompatibility Thesis is false. Given that team reasoning is an important form of reasoning we

sometimes engage in when cooperation is called for, this result strengthens the case for the content approach.

Acknowledgements

Thanks to Michael Bratman, Stephanie Collins, Julian Davis, Raul Hakli, Rory Harder, Yair Levy, Bhaskarjit Neog, Björn Petersson, Abe Roth, Frans Svensson and participants of the Practical Philosophy research seminar at University of Gothenburg for comments on earlier drafts. John Horden proofread a version of the chapter, and I improved the grammar and style of a few passages in a later version with the help of Microsoft Copilot. My research was funded by the Lund Gothenburg Responsibility Project (PI: Paul Russell), which in turn was funded by the Swedish Research Council, an HCAS-RJ fellowship at the Helsinki Collegium of Advanced Studies, University of Helsinki, 2025–2026, and the project Commitment and Self-Evaluation: Individual and Collective (PIs: Arto Laitinen and Lilian O'Brien), funded by the Research Council of Finland.

Notes

1. Insofar as such obligations and entitlements seem unintelligible on the reductive content approach that I take (Gilbert 2022), my defence of it is incomplete (although see Bratman 2014: ch. 5).⏎
2. For a classic defence of a reductive causal theory of action, see Bishop (1989). For discussions of joint intentional action from a different, Anscombian starting point, see e.g. Stoutland (2008), Laurence (2011) and Satne (2024). I set up a challenge for such Anscombian views in Blomberg (2018).⏎
3. The term 'we-intention' was introduced by Wilfred Sellars (1980: 97, paragraph 161). Michael Bratman suggests that his intention condition (see Section 3) points to 'reductive sufficient conditions for something analogous to a Searle-type

we-intention.' (2014: 154) Kirk Ludwig (2016: ch. 13) gives a reductive account of 'we-intention'.⏎

4. Marija Jankovic and Kirk Ludwig (2026: 113), who endorse the content approach, remark: 'Plausibly we do often engage in team reasoning and form as a result shared intentions [matching we-intentions]'. But they do not explain in detail how this is supposed to work. (Unlike Michael Bacharach, Ludwig (2017) takes team reasoning to be functionally equivalent, roughly, to what I in Section 4 call best-reply reasoning of team benefactors) Others have suggested or argued that intentions that they *J* together can prompt team reasoning (Gold and Sugden 2007: 117, 136–7; Ludwig 2017: 13–16; cf. Hakli and Mäkelä 2017: 134, 137–8), or that a we-intention is simply a team reasoner's ordinary intention to merely perform her role in the *J*-ing, rather than an intention that they *J* together (Kutz 2000: 7–10; Gold and Sugden 2007; Pacherie 2013). For critical discussion of the latter idea, see Bratman (2014: 181–2 n. 19) and Jankovic and Ludwig (2026: ch. 9).⏎
5. Raimo Tuomela (2007) also takes we-intentions to be irreducible to ordinary intentions. Unlike Searle, Tuomela is a non-individualist: a we-intention is a relational state that ontologically depends on the participants' 'joint intention' that is 'nonaggregative and emergent' (2007: 131).⏎
6. There are alternatives to the mode approach that allow one to retain the own-action condition, though they are *prima facie* less plausible. Stephanie Collins (2019), who is sympathetic to the own-action condition (104–6), takes a *group agent approach* (44–45). On this approach, the snowboarding friends who are riding down the slope jointly intentionally—unlike the strangers each snowboarding on their own—would constitute a group agent with its own intention to ride down the slope.

While an organised institution, such as a corporation, may be an agent with its own intention, the suggestion that the snowboarding friends make up a group agent whose intention rationalises and causally guides the friends' joint activity is difficult to accept. On a *shared-token-intention approach*, the snowboarding friends—unlike the strangers—would instead literally share a token intention that rationalises and causally guides their joint activity. However, several minds cannot share a single token mental state without fusing into one mind (*pace* Velleman 1997: 38; Wilby 2012; Schmid 2023: ch. 2). (While snowboarding together with friends is wonderful, it does not involve a literal fusion of minds.) Perhaps, however, a shared token intention could be a *non-mental* intention (Velleman 1997: 37–38). Alternatively, a 'shared intention' could be merely a functional analogue of a token intention (Bratman 2014: 10, 27, 86; Roy and Schwenkenbecher 2021). For those who accept the own-action condition, the challenge would be to demonstrate that such a functional analogue need not involve participants' intentions that they *J* together.⏎

7. Ludwig (2016: 102–4) argues that intentions that they *J* together are always really intentions *to bring it about* that they *J* together (for critical discussion, see Núñez 2022: 489–492). Ludwig thus accepts the own-action condition.⏎
8. For an internal critique of Bratman's way of avoiding circularity, see Petersson (2007). For a response, see Bratman (2014: 93–95). For an external critique, see Kietzmann (2018).⏎
9. I call this condition the 'doxastic single end condition' and give it a different formulation in Blomberg (2016a). The difference is merely terminological.⏎
10. Even if intending were subject to a strong belief condition, according to which I could intend to φ only if I believed I

would φ, the participants could have matching we-intentions without this fact being common knowledge between them and without being irrational. Consider the following case (Blomberg 2016b): Hector and Celia each intend that they build a block tower together; they each intend that they do this by way of each of these intentions, and by way of meshing sub-plans. They each also believe that they both have these intentions. Furthermore, they each believe that the end of them building a block tower together is shared, such that they each intend it. However, Hector falsely (but perhaps justifiably) believes that Celia falsely assumes that he merely intends to cover the top surface of each of her blocks. Despite this false belief, their intentions could still successfully coordinate their construction of the block tower, and each of them could be fully responsible for making it the case that they build it together. There is no reason, I contend, that such an activity should not count as jointly intentional.⏎

11. 'This assumption claims that [the extent of action whose qualities (goodness, rightness) provide reasons for or against performing some action *A*] always matches, and is determined by, the unit of agency' (Woodard 2011: 263).⏎
12. As far as I know, Bacharach's hypothesis that strong interdependence triggers group identification has not been experimentally tested.⏎
13. My discussion of Petersson's (2017) argument draws on, and further develops, my discussion in Blomberg (2023).⏎
14. Following François Recanati, it may be useful to distinguish between the 'complete' and the 'explicit' content of an intentional state, where 'the complete content of the state is jointly determined by its [explicit] content and by its mode' (2007: 22). Those aspects of the complete content that are

determined by the mode are ‘only implicitly represented’ (2007: 22).⏎

15. I am not aware that Bratman has either rejected or accepted team reasoning as a valid form of practical reasoning.⏎

References

Asarnow, Samuel. 2020. ‘Shared Agency Without Shared Intention’, *The Philosophical Quarterly*, 70(281): 665–688.⏎

Bacharach, Michael. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, N.J.: Princeton University Press.

Bishop, John. 1989. *Natural Agency: An Essay on the Causal Theory of Action*. Cambridge: Cambridge University Press.⏎

Blomberg, Olle. 2011. ‘Socially Extended Intentions-in-Action’, *Review of Philosophy and Psychology*, 2(2): 335–53.⏎

Blomberg, Olle. 2016a. ‘Shared Intention and the Doxastic Single End Condition’, *Philosophical Studies*, 173(2): 351–72.⏎

Blomberg, Olle. 2016b. ‘Common Knowledge and Reductionism about Shared Agency’, *Australasian Journal of Philosophy*, 94(2): 315–26.⏎

Blomberg, Olle. 2018. ‘Practical Knowledge and Acting Together’, in J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos and Duncan Pritchard (eds), *Socially Extended Epistemology*, pp. 87–111. Oxford: Oxford University Press.⏎

Blomberg, Olle. 2021. ‘Intentional Cooperation and Acting as Part of a Single Body’, *Mind and Language*, 36(2): 264–84.⏎
Blomberg, Olle. 2023. ‘Team Reasoning, Mode, and Content’, in Andrés G. Garcia, Mattias Gunnemyr and Jakob Werkmäster (eds), *Value, Morality & Social Reality: Essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen*, pp. 39–54. Lund: Department of Philosophy, Lund University.⏎
Blomberg, Olle and Frank Hindriks. 2020. ‘Collective Responsibility and Acting Together’, in Saba Bazargan-Forward and Deborah Tollefsen (eds), *The Routledge Handbook of Collective Responsibility*, pp. 142–54. Abingdon: Routledge.⏎
Blomberg, Olle and Erik Malmqvist. 2024. ‘Cooperative Activity, Shared Intention, and Exploitation’, *Ethics*, 134(3): 387–401.⏎
Blomberg, Olle and Björn Petersson. 2024. ‘Team Reasoning and Collective Moral Obligation’, *Social Theory and Practice*, 50(3): 483–516.⏎
Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press. Reissued by CSLI Publications 1999.⏎
Bratman, Michael E. 1997. ‘I Intend that We *J*’, in Ghita Holmström-Hintikka and Raimo Tuomela (eds), *Contemporary Action Theory, Volume 2: Social Action*, pp. 49–63. Dordrecht: Kluwer.⏎

Bratman, Michael E. 2014. *Shared Agency: A Planning Theory of Acting Together*. New York, NY: Oxford University Press.⏎

Browne, Katharine. 2018. ‘Why Should We Team Reason?’, *Economics and Philosophy*, 34(2): 185–98.⏎

Butterfill, Stephen A. 2015. ‘Planning for Collective Agency’, in Catrin Misselhorn (ed.), *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, pp. 149–68. Cham: Springer Verlag.⏎

Butterfill, Stephen. 2016. ‘Joint Action: A Minimalist Approach’, in Julian Kiverstein (ed.), *The Routledge Handbook of Philosophy of the Social Mind*, pp. 357–69. Abingdon: Routledge.⏎

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.⏎

Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.⏎

Colman, Andrew M. and Natalie Gold. 2018. ‘Team Reasoning: Solving the Puzzle of Coordination’, *Psychonomic Bulletin & Review*, 25(5): 1770–83.⏎

Fodor, Jerry and Ernest Lepore. 1992. *Holism: A Shopper’s Guide*. Cambridge, MA: Blackwell.⏎

Gilbert, Margaret. 2022. ‘A Simple Theory of Acting Together’, *Journal of the American Philosophical Association*, 8(3): 399–408.⏎

Gold, Natalie and Robert Sugden. 2007. ‘Collective Intentions and Team Agency’, *Journal of Philosophy*, CIV(3): 109–37.⏎

Hakli, Raul and Pekka Mäkelä. 2017. ‘Planning in the We-mode’, in Gerhard Preyer and Georg Peter (eds), *Social Ontology and Collective Intentionality: Critical Essays on the Philosophy of Raimo Tuomela with His Responses*, pp. 117–40. Cham: Springer.⏎

Hakli, Raul, Kaarlo Miller and Raimo Tuomela. 2010. ‘Two Kinds of We-Reasoning’, *Economics and Philosophy*, 26(3): 291–320.⏎

Hindriks, Frank. 2012. ‘Team Reasoning and Group Identification’, *Rationality and Society*, 24(2): 198–220.⏎

Hogg, Michael A., Dominic Abrams and Marilynn B. Brewer. 2017. ‘Social Identity: The Role of Self in Group Processes and Intergroup Relations’, *Group Processes & Intergroup Relations*, 20(5): 570–81.⏎

Jankovic, Marija and Kirk Ludwig. 2026. *An Introduction to Collective Intentionality: In Action, Thought, and Society*. Abingdon: Routledge.⏎

Kietzmann, Christian. 2018. ‘Bratman on Shared Intention’, *SATS: Northern European Journal of Philosophy*, 19(2): 161–81.⏎

Kutz, Christopher. 2000. ‘Acting Together’, *Philosophy and Phenomenological Research*, 61(1): 1–31.⏎

Laurence, Ben. 2011. ‘An Anscombian Approach to Collective Action’, in Anton Ford, Jennifer Hornsby and Frederick

Stoutland (eds), *Essays on Anscombe's Intention*, pp. 270–94. Cambridge, MA: Harvard University Press.⏎

Ludwig, Kirk. 2016. *From Individual to Plural Agency. Collective action: Volume 1*. Oxford: Oxford University Press.⏎

Ludwig, Kirk. 2017. 'Methodological Individualism, The We-mode, and Team Reasoning', in Gerhard Preyer and Georg Peter (eds), *Social Ontology and Collective Intentionality: Critical Essays on the Philosophy of Raimo Tuomela with His Responses*, pp. 3–18. Cham: Springer.⏎

Mathiesen, Kay. 2002. 'Searle, Collective Intentions, and Individualism', in Georg Meggle (ed.), *Social Facts and Collective Intentionality*, pp. 185–204. Frankfurt am Main: Hänsel-Hohenhausen.⏎

Núñez, Carlos. 2022. 'Intending Recalcitrant Social Ends', *Erkenntnis*, 87(2): 477–98.⏎

Pacherie, Elisabeth. 2013. 'Intentional Joint Agency: Shared Intention Lite', *Synthese*, 190(10): 1817–39.⏎

Petersson, Björn. 2007. 'Collectivity and Circularity', *Journal of Philosophy*, 104(3): 138–56.⏎

Petersson, Björn. 2017. 'Team Reasoning and Collective Intentionality', *Review of Philosophy and Psychology*, 8(2): 199–218.⏎

Pettit, Philip. 1995. 'The Virtual Reality of *Homo Economicus*', *The Monist*, 78(3): 308–29⏎

Recanati, François. 2007. *Perspectival Thought: A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.⏎

Roth, Abe. 2020. ‘Proprietary Reasons and Joint Action’, in Anika Fiebich (ed.), *Minimal Cooperation and Shared Agency*, pp. 169–80. Cham: Springer.⏎

Roth, Abraham Sesshu. 2014. ‘Team Reasoning and Shared Intention’, in Anita Konzelmann Ziv and Hans Bernhard Schmid (eds), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology*, pp. 279–95. Dordrecht: Springer.⏎

Roy, Olivier and Anne Schwenkenbecher. 2021. ‘Shared Intentions, Loose Groups, and Pooled Knowledge’, *Synthese*, 198(5): 4523–41.⏎

Satne, Glenda Lucila. 2024. ‘Practical Knowledge and Shared Agency: Pluralizing the Anscombean View’, *Inquiry*, 67(4): 1018–45.⏎

Schmid, Hans Bernhard. 2023. *We, Together: The Social Ontology of Us*. New York, NY: Oxford University Press.⏎

Schmitz, Michael. 2017. ‘What Is a Mode Account of Collective Intentionality?’ in Gerhard Preyer and Georg Peter (eds), *Social Ontology and Collective Intentionality: Critical Essays on the Philosophy of Raimo Tuomela with His Responses*, pp. 37–70. Cham: Springer.

Schwenkenbecher, Anne. 2021. *Getting Our Act Together: A Theory of Collective Moral Obligations*. New York, NY: Routledge.⏎

Searle, John R. 1990. ‘Collective Intentions and Actions’, in Philip R. Cohen, Jerry Morgan and Martha E. Pollack (eds),

Intentions in Communication, pp. 401–15. Cambridge, MA: MIT Press.⏎

Searle, John R. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.⏎

Sellars, Wilfrid. 1980. ‘On Reasoning about Values’, *American Philosophical Quarterly*, 17(2): 81–101.⏎

Stoutland, Frederick. 2008. ‘The Ontology of Social Agency’, *Analyse & Kritik*, 30(2): 533–51.⏎

Sugden, Robert. 2000. ‘Team Preferences’, *Economics and Philosophy*, 16(2): 175–204.⏎

Tuomela, Raimo. 2007. *The Philosophy of Sociality: The Shared Point of View*. New York, NY: Oxford University Press.⏎

Velleman, David. 1997. ‘How To Share An Intention’, *Philosophy and Phenomenological Research*, 57(1): 29–50.⏎

Wilby, Michael. 2012. ‘Subject, Mode and Content in “We-Intentions”’, *Phenomenology and Mind*, 2: 93–106.⏎

Woodard, Christopher. 2011. ‘Rationality and the Unit of Action’, *Review of Philosophy and Psychology*, 2(2): 261–77.⏎

OceanofPDF.com

12 The Second-person Perspective on the First-person Plural

Bijoy H. Boruah

DOI: 10.4324/9781003612827-13

The ontology of the social is problematic because of the way its plural constituents are supposed to be bonded into a single integral whole, a *single* but not 'singular' reality. A social reality incorporates multiplicity into its ontological fold, and it looks as though an emergent unity 'absorbs' the singular identities of the constituent members of the social whole. However, singular identities are diehard entities, and therefore the metaphor of absorption used in accounting for the emergence of a social unit as a *sui generis* reality has relentlessly been subjected to various demands for philosophical articulation. Individualist ontology, in the sense of reckoning the individual person as ontologically primary in the constitution of social reality, has a tremendous intuitive hold in the philosophical mind against granting ontological primacy to the social whole. That is precisely why social ontology has figured as a problem of philosophical concern. But the philosophical concern is also made urgent by the apparent belief that the social is ontologically no less a diehard entity. Individualism is challenged in an insistent contest for the ontological autonomy of the social.

It seems, on the one hand, that the study of our social existence, in its most basic form, requires us to start with the question of how it is possible to leap out of one's individual subjectivity. On this theoretical approach, one must begin with the I, the individual subject, and proceed outward in a bid to identify the conditions under which one can recognise or know another I. On the other hand, there is a pervasive and anonymous form of sociality that constitutes the very rhythm of human existence. The theorising on this approach is to start with the social and work inward,

finding the conditions under which individual consciousness is situated within a social context.

In this chapter, I shall regulate my thoughts on social ontology in terms of the intuitive appeal of individualist ontology, or the ontological primacy of the individual, on the fundamental ground that consciousness is structurally sourced in an individual conscious subject. The first-person perspective of consciousness is the perspective of the first-person *singular,* the *I* or self that forms the locus of individual subjectivity and personhood. There is no collective or plural subject as a *primary* source or locus of consciousness, no matter how naturally we are accustomed to accepting the notion of 'collective consciousness' in discourses about life and society. On this view, therefore, the theoretical approach is that of working from within —the matrix of individual consciousness—to move outwards for explaining the possibility of the social.

1 Two Foci of the Ellipsis of Consciousness: *I*-Perspective and *We*-Perspective

It is at this juncture that the relationship between consciousness and subjectivity poses a problem. Is the first-person perspective of consciousness ontologically divided into two distinct modes—*I*-consciousness and *We*-consciousness—corresponding to the grammatical dichotomy of first-person singular and first-person plural in the pronominal framing of speech acts or sentential expressions? Is the formal grammatical separation of first-personal pronouns into the singular and the plural categories metaphysically underpinned by a parallel separation of individual and social consciousness as *equally primordial loci* of conscious subjectivity?

This problem of comparing and connecting the formal structure of grammar with its substantive ontological counterparts arises in the prevailing debate on the theory of collective intentionality, which is underpinned by the centrality of the nature of the *We*. Whilst the first-person singular or *I* is the subject of consciousness, grammar offers the first-person plural or *we* as a parallel subject to construct sentences of collective thought or action. It is the parallel existence of *We*-subjectivity that engenders the

persuasive belief in the irreducible ontology of collective or social consciousness. There are two possibilities that open up in regard to this problem of We-subjectivity.

One is the ontological, *subject*-based reading, which refers to the existence of an irreducible *plural* subject as the locus of we-subjectivity. There is a *single* (but not singular) ontological bearer, the *We*-subject, which obtains in virtue of how its constituent individual members relate with one another. This single plural subject or *We* is the subject of a corresponding single collective intentional state, a *We*-intention, which is entertained by all the members of a collective. This theory of a plural subject of a We-intentional state is championed by Margaret Gilbert in various works (1989, 2003, 2007).

The other theory, professed in various works by John Searle (1990, 1995, 2010), advocates a *mode*-based interpretation in that it doesn't commit itself to the existence of any kind of subject other than the first-person singular *I* correspondingly to the existence of collective intentional states. The individual subject is either in the singular 'I intend' *mode* of entertaining singular I-intentional states, or in the plural 'we intend' mode of entertaining collective we-intentional states. This situation obtains in virtue of a collection of individuals, tokening a type-identical intentional state like 'we are walking to the beach' or 'we have allegiance to the authority of the Queen'.

The backdrop of the discussion of this chapter is whether We-sociality is an innate and primitive disposition in individual consciousness, or it is actualised through a mediated process of socialisation that the individual subject enters into. There are two versions of a prevailing view on this issue. In one view, there are two separate capabilities—*I*-intentionality and *We*-intentionality—between which the individual subject can switch, and thereby realise the ontological condition of We-sociality. An alternative to this is the view that, rather than being two separate capabilities, the *I* and the *We* are two divergent focal points originating in the same first-person perspective of consciousness (Carminada 2016: 292).

I describe both versions of this view as *immanentist* in that they take we-sociality to be innate in the first-person singular subject. In this view, the *We* is not an *explanandum* which can be grasped only through the completion of a process that constitutes the *explanans*. In other words, the immanentist,

or innatist, view discounts the necessity of explaining how the first-person singular subject needs to 'reach out' beyond its singular, *intra*-subjective ontological boundary to a wider space of inter-subjectively plural formation. We-sociality then becomes an irreducible and primordial feature of the individual human subject, the first-person singular, who is already equipped with the we-social frame of mind. Since we-ness is constitutively embedded in I-ness, and individuality is already enmeshed in sociality, it then follows that there is no ontological gap between first-person singular and first-person plural to be bridged by a transition from subjectivity to inter-subjectivity.

The formalist discourse of grammar is understandably silent about whether the grammatical rule for the use of the first-person pronoun in its singular and plural incarnations—as *I* and as *We*—is indicative of the ontological primacy of the one over the other. Suppose we interpret *I* and *We* on the model of an ellipsis that has two foci necessarily interlinked with one another (Carminada 2016: 91-92). It would then mean that they are two foci of a common first-person perspective of consciousness. That is, consciousness's first-personal orientation is either I-perspectival or we-perspectival, and these two perspectival orientations—singular and plural—are already mutually implied in one another.

The 'elliptical' interpretation of the I/We difference neutralises the ontological determination of *being I*, understood as the reality of the first-person singular subject of consciousness on the one hand, and *being We* that refers to the reality of the first-person plural subject on the other. This interpretation has the advantage of liquidating the need for a substantive metaphysical construal of the subject of consciousness in either singular or plural incarnation as some kind of an *entity*. *I* and *We* are not metaphysically enigmatic subjective entities but only *perspectives*, and perspectives have no being of their own. Construing the subjectivity of consciousness in 'perspectival' terms thus frees us from ontological reification of the conscious subject.

It must be conceded, however, that a perspective is an indexical which has to be indexed to a particular position or point in space and time. Take the cases of *here* and *now*, the two familiar examples of spatial and temporal indexicals. If I say that 'I am here right now' when, at this very time (say 25 December 2024), I am attending a Christmas event in Vienna,

it means that ‘here’ is indexed to Vienna and ‘now’ to 25 December 2024. Likewise, if I-perspective and we-perspective as first-personal perspectival orientations of consciousness must be indexed to determinate positions or points in the space of consciousness. Since these perspectives of consciousness cannot float free as anonymous orientations, they must in some sense be ‘owned’ (by particular conscious subjects) to determine the source or locus of the perspectival orientation.

Consciousness’s perspectival orientation is to be understood in terms of the intentionality or world-directed nature of consciousness. The subject of consciousness is aware of the world from its (subjective) perspective, which means that first-person subjectivity—the *I*-perspective—structurally determines consciousness’s relation to the world. The I-perspectival orientation is clearly anchored in the individual conscious subject; the individual subject ‘owns’ the perspectival orientation it has to the world. So, the I-perspective of consciousness must be indexed to the individual subject. The subject or I cannot therefore be reduced to a mere perspective. In other words, *being I* must be a reality to provide anchorage to the first-personal perspectival orientation of consciousness.

What, then, is the we-perspectival orientation of consciousness anchored in? What is the locus of we-consciousness or first-person plural subjectivity? The difficulty of answering this question is also the fulcrum of social ontology as a philosophical inquiry into social reality. If we are to grant *being we* as a reality and the plural counterpart to the reality of *being I*, we confront the metaphysically vexing question of committing to the existence of a *plural subject* of consciousness—an autonomous ‘collective subject’ parallel to the existence of the autonomous ‘singular subject’ in the same orbit of consciousness. The problem is: how does one map the *grammatical* first-person plural onto the *ontological* architecture of consciousness without having to posit a communal or collective subject in the generically homogeneous space of consciousness?

In this chapter, I consider the difficulty of postulating a communal subject as an ontological correlate of the first-person plural pronoun to be indicative of an asymmetry between grammar and ontology. The first-person perspective systemically bifurcates into the I-perspective and the we-perspective as formal, grammatical moves in the communicative practice of natural language. But it is far from obvious that the orbit of

consciousness analogously divides into the I-subject and the we-subject conceived as distinct representational loci of subjectivity. What exactly is the existential configuration of the so-called *we*-subject?

The above discussion foregrounds the development of my idea on we-sociality in this chapter. I argue that *We*-sociality is not just a matter of intersubjectivity, of relations between individual subjects like *you* and *me*, interacting in a shared space of plurality. Rather, this intersubjective space of interactive involvement must be seen as an *accomplishment* of collectivity on the part of engaged individual members. That is, individual subjects of consciousness transcend their individuality by transitioning to an intersubjective relational process that culminates in the formation of we-sociality, which amounts to attaining the consciousness of being *us*. My position understandably differs from the immanentist conception, which regards we-sociality as innate in the I. In my view, we-sociality in the multitudinous sense of first-personal plurality is to be explained as a phenomenon mediated by an initial process of minimally intersubjective second-personal engagement.

As I said earlier, I follow the intuitive appeal of the belief that the first-person perspective is intrinsically individualist because the conscious subject to which the stream of consciousness is anchored is an individual subject. Given the primacy or primordiality of the first-person singular as a fundamental perspective of consciousness—that is, the perspective of *I*-consciousness—the reality of consciousness that is anchored to the first-person plural perspective—that is, the perspective of *we*-consciousness—would seem to be an *emergent* reality and a function of I-consciousness. At this point, I shall assume that the emergentist interpretation of we-consciousness is plausible. The question then is how to explain the implied transition from the singular first-person perspective of being the self-aware subject to the plural first-person perspective of becoming a *social* subject, a locus of consciousness that integrates multiple subjects of singular self-awareness into a single unit, a collective. How does one move from the individually subjective interiority of I-consciousness and work one’s way outwards to a shared or intersubjective state of we-consciousness?

The ‘aggregation’ interpretation of sociality is discredited because the mere aggregation of a plurality of conscious individual subjects would not yield a proper collective *unit* to constitute a unified centre of multiple

subjects of consciousness. The relation of aggregation of multiple distinct individual subjects of consciousness by itself does not provide the cementing force needed for the bonding of a social whole—that is, the phenomenological unity of we-consciousness (Tollefsen 2004). So the relationality in question has to be characterised by other properties and processes than aggregation. It is also no good to resort to metaphors of ‘merger’ and ‘fusion’ of distinct first-person singular perspectives in multiples. Singular perspectives of I-consciousness in multiples do not blend into a collective we-consciousness because distinct perspectives in their singularity must maintain or preserve their distinctness if the *sense* of being a part of the collective is essential to the reality of we-consciousness. There is no fusion of multiple centres of first-person perspectives in their singularity or distinctness into a single super-consciousness which absorbs a plurality of centres of consciousness.

2 We-Sociality Thin and Thick: From the Minimal I-You to the Multitudinous I-We

What we therefore might want to explore is the idea of some kind of a *union without fusion*, a whole of interrelated parts, a single *we*-subject constituted of multiple interrelated *I*-subjects. The crux of the problem in this idea is how to describe the nature of this interrelation without making the collective subject a mere collection of individual subjects. But the most important part of this problem is to realise that we are talking about *interpersonal* relatedness, not merely between one individual person or I-subject and another fellow individual I-subject, but an individual subject or person and a multitude of other individual I-subjects. It is a case of *one*-with-*many* interpersonal communion that needs to be the object of analysis and articulation. Of course, it takes just *two* to form a We; but an example of interpersonal connectedness within a dyadic constellation is much too impoverished or thin to constitute the ‘full-blooded’ sociality of We. We need an interpretation and understanding of ‘we’ in its robust or thick sense of a social whole, which emerges in interpersonal connectedness within a multitude of persons.

Is this difference between the robust sense of *we* and the thin version of it merely a matter of one being *larger* than the other in terms of interpersonal connectedness? It looks as though the thin version represents the first-person plural as an indicator of mini-sociality, the smallest unit of the collective we-subject consisting of two distinct I-subjects in their interrelatedness. Formally and grammatically speaking, this is a true account of minimal sociality in the sense of interpersonal 'bonding' required by we-sociality. Now, we confront the question of whether the interpersonal bonding involved in thick we-sociality has to be different in nature from that of thin we-sociality. In case there is a difference in the nature of bonding in the two versions, the further question to confront is: whether the transition from thin we-sociality to thick we-sociality implies the formation of a *new* kind of interpersonal bonding altogether, which is characteristically discontinuous with the interpersonal bonding involved in the thin case. Is interpersonal bonding occurring within a multitude of persons *qualitatively* different from that of bonding within a dyadic relation?

First, we need to address the nature of thin we-sociality or dyadic interpersonal connectedness: that is, the second-person singular perspective of I-Thou relation, in which two I-subjects address one another by adopting the You-stance and thereby participate in a dialogical space of reciprocity. I shall describe this dialogical space as the we-subjective space of dialogical reciprocity. Two distinct I-subjects—You and I—enter into an addressive relational nexus, thereby generating a minimally extensive, or minimally plural, we-subject within a circumscribed inter-subjective communicative space.

Imagine the beginning of the actual human world as populated only by two persons, Adam and Eve. In that imaginary situation, the first addressive-communicative scenario occurs between two first-personal perspectives: that of one I-subject, Adam, and of another I-subject, Eve. There occurs the first ever I-Thou relational nexus of Adam and Eve, and out of this nexus there emerges the primeval we-subjectivity, the very first we-subject as a kind of plural subject. Two singular subjects in their distinctness come together through the second-personal perspective, or the *you*-stance. The alchemy of this togetherness of two individual I-subjects culminates in the unique formation of a minimally plural (dyadic)—hence

minimally social—we-subject. Whilst Adam and Eve are two independent centres of the first-person perspective, their engagement in the second-person perspective allows the two separate I-subjects to become a single we-subject.

Suppose that Adam and Eve are, in the imagined human world, the only occupants of an inter-subjective I-thou relational space. This would be the maximally circumscribed space of minimal plurality, therefore of minimal we-sociality. This fictionalised world scenario presents the picture of the emergence of the first we-subject of minimal sociality. Since this we-set consists only of two members who relate with one another by adopting the reciprocal you-stance, this inter-personal relationship of we-formation lacks the plenitude of multitudinous sociality. Here the ‘we’ amounts to no more than ‘I and you’ or ‘you and I’. The intersubjective space of interpersonal betweenness is fully exhausted by two singular subjects as the only occupants. If it is a social world, its sociality is, as it were, on the verge of vanishing or collapsing into individuality. Any time one of the two subjects drops out of the I-thou relation, or simply disappears by demise, the minimal sociality collapses into ontological individuality. Minimal sociality is therefore precarious sociality due to its utter lack of plenitude. But extended, thick, or robust sociality has an enduring character because of multitudinous membership in its constitution. If one or two members drop out or disappear from the we-set, we-sociality can still survive the diminishing effect on it.

Now, imagine further that, unbeknownst to Adam and Eve, the world was already populated with a very large number of other human persons, and they then all met to get to know one another. This would obviously create the possibility of an expanded we-sociality, and a robust inter-subjective, inter-personal I-we relationship. What changes in this new scenario, formally or structurally, is that the I becomes part of a much bigger, one-*many* relational nexus than the earlier situation, in which the I was part of a one-*one*, I-thou relational nexus. Is this transition from narrow *I-thou* sociality to wide *I-we* sociality a radical shift to a completely new form of sociality, a new bonding of collective subjectivity transcending earlier, and narrower, inter-subjective self-awareness that characterises the I-thou relation?

In articulating the second-person perspective in terms of a specific kind of interpersonal 'engagement', Dan Zahavi writes:

> For me to relate to and address another as a you is to relate to someone, an I, who is in turn related to me as a you. Second-person engagement is a subject–subject (you-me) relation where I am aware of and directed at the other and, at the same time, implicitly aware of myself in the accusative, as attended to or addressed by the other. Second person engagement consequently involves not merely an awareness of the other, but also and at the same time, a form of *interpersonal self-consciousness*.
>
> (Zahavi 2019: 255, emphasis added)

Emphasis is laid here on reciprocity characterised by the phenomenology of interpersonal self-awareness in taking the second-personal stance. Zahavi connects his interpretation with that of Husserl, who describes this reciprocal interaction between the two I-subjects, or between the I and the other person, as a form of 'communicative intertwinement' (Zahavi 2019: 255).

What makes this communicative intertwinement possible for Husserl is the fact that this addressive-communicative exchange involves an internalisation of the other's perspective on myself. Zahavi views this internalisation as one's attaining 'a socially mediated externalised self-apprehension' (Zahavi 2019: 255) and considers it to be crucial for the formation of we-sociality through we-identity. It is by adopting the view the other has of me, or internalising the other's perspective on me, that, according to Husserl, '*I fit myself into the family of man*', thereby creating the 'possibility for the unity of this "family" [of humankind]', so that, and only after that, I become, 'in the proper sense, an Ego over against an other and can then say "we"' (Husserl 1989: 254).

Husserl says here something ontologically significant about one's being a *person* in terms of some kind of achievement by transcendence. It is a kind of ontic conversion of oneself as a singular self-enclosed I into a

‘proper’ Ego or personal self, through its being intertwined with others and acquiring we-identity. It is only when a singular I is essentially integrated into the human we-family—that is, the I becoming *one of us*—does the I become a proper personal self. This does not mean, however, that the first-person plural or we-subject is an enlargement of the first-person singular or I-subject. The we is not equivalent to an enlarged I in virtue of the singular I’s mysterious fusion with a plurality or multitude of other individual selves, thereby becoming an inflated subject of plural first-personal perspectives. Even though it is true that there is a sense of merger and unifying integration of the singular *I* in the plural *we*, this coalescence of independent subjects of consciousness never equals an ontologically inflated inseparable macro-subject that absorbs the distinct first-personal micro-subjects. Rather, we-formation is predicated on the preservation of plurality of individual subjects. The presence of differentiated or individual centres of consciousness is a condition of possibility of we-identity and sociality.

3 Bridging over Betweenness

What the unifying integrity of we-sociality or we-identity involves is the *bridging* of the gap between the I and others rather than *erasing* differentiation to create an undifferentiated fusional we-awareness (Zahavi 2020: 19–24). The gaps are bridged between an I and any other I-subject(s), and it is this *bridge over betweenness*, as it were, which enables individual I-subjects to leap beyond their first-personal self-awareness into a we-identity of interpersonal self-consciousness.

However, this leap or transcendence is not a reflective or conscious act, on the part of the I-subject, of being inducted into the human condition of we-sociality from a prior, pre-social stage of primordial I-individuality. The individual self is naturally a social self as well in the sense that every individual I is in some way or other an integral part of some *we*. So, the gap between the individual and the social is both real and yet always bridged by a natural human disposition for interpersonal connectivity.

It is at this juncture that the matter of *explaining* we-identity and group-membership in terms of interpersonal relations between individual I-

subjects becomes critical. If the formation of we-sociality is not predicated on any conscious or reflective engagement in connecting oneself with others in we-groups, does it mean that the 'we' does not originate in any kind of agreement or communicative reciprocity, or joint action of mutual relation? Are individual I-subjects fundamentally embedded and organically integrated into the we-social environment?

What readily comes to mind is Heidegger's thoughts, in which we find an affirmative answer to these questions: 'So far as Dasein is at all, it has Being-with-one-another as its kind of Being' (Heidegger 1962: 163). One cannot *be* the kind of human being one is in the world without the property of *being with* others. Being-oneself is essentially being oneself-*with*-others, such that sociality is built into the very structure of individual human existence. An elaborate description of this essential sociality of individual human existence is to be found in a later work by Heidegger:

> Dasein is essentially *being-with* others *as being-among* intraworldly beings. As being-in-the-world it is never first merely being among things extant within the world, then subsequently to uncover other human beings as also being among them. Instead, as being-in-the-world it is being with others, apart from whether and how others are factically there with it themselves… . Put otherwise, being-in-the-world is with equal originality both being-with and being-among.
>
> (Heidegger 1982: 278)

The 'others' that one is essentially said to be *with* importantly deserves a clarification that Heidegger provides:

> By 'Others' we do not mean everyone else but me—those over against whom the 'I' stands out. They are rather those from whom, for the most part, one does not distinguish oneself—those among whom one is too… By reason of this *with-like*

> Being-in-the-world, the world is always the one that I share with Others.
>
> (Heidegger 1962: 154–5)

What is clear here is that the ‘others’ that I am with are not *all* the other people I stand over against in this human world. Rather, they are those others with whom I *share* a common life, such as being a member of a group or community, with a sense of *belonging* or we-identity. And the world that I share with others in a regular way—the shared world of everyday Dasein—is a *lifeworld* rather than an objectively, and social-scientifically, described human world—that is, the larger world inhabited by all the other fellow humans, who are designated by the third-person plural ‘they’ (rather than its first-person plural counterpart ‘us’).

The above clarification yields the further important point that there is a phenomenon of *subjective internalization* of objectively prevalent *others*, or other fellow human beings, in the larger human world. Becoming a member of a we-group, and experiencing we-identity with the members, implies that one internalises others into the group one belongs to. Others then become *co-personal* members of the group or community that I reckon as *us*. There is a ‘first-personal affinity’ to others who are experienced by me as one of *us* rather than reckoned as one of *them*. A social intimacy (feeling of oneness) that grows from first-personal affinity to others is a defining feature of we-identity. By contrast, others-as-*they* remain a kind of anonymous social reality to which I am more or less indifferent, or relate to them from what might be called a ‘third-personal’ distance. It is others-as-*us* of everyday Dasein and lifeworld, from whom I do *not* stand away for the most part of my life. By contrast, it is the larger humanity of fellow human members, designated as others-as-*they*, over whom I *do* stand out in a more-or-less neutral or non-committal standing-relation.

If the phenomenon of *subjective internalization* is characteristically implied by my relationship with others-as-*us*, then the formation of we-identity and we-sociality necessarily involves adopting a first-personal stance towards others. Even if I am always already in the midst of a social environment—that is, my *being* in the world is my *being with* others in the

world—I become we-conscious of the world only when I find a first-personally acclimatised accommodation, so to speak, in a larger, third-personal space of humanity. The space of we-sociality is therefore *first-personally circumscribed* out of a largely third-personal space of social reality full of others, even though others like me or us. Indeed, the social world is much wider than what the net of we-sociality can spread over in virtue of first-personal acclimatisation in the formation of we-identity.

The metaphor of 'first-personal acclimatization' suggests a process that the I goes through—an experience of interpersonal self-consciousness in relation to other I-subjects. Nothing is more salient as an example of interpersonal self-consciousness than the experience of I-Thou communicative relationship. Not just salience; it is also a fundamental form of intersubjectivity in that two singular first-person perspectives become mirror images of each other. This is what transpires in the 'second-person engagement' that Zahavi alludes to in highlighting subjective reciprocity. The I (first person) reaches out to the otherness of the second person in the same wave-length of subjectivity—that is, as another centre or locus of the first-person perspective. Thus, the you-stance creates a common space of subjectivity shared by two distinct subjects, where their individual first-person perspectives interpenetrate and intertwine to generate a primary we-consciousness, which is the most basic form of interpersonal self-consciousness.

Martin Buber is an indispensable source to invoke in understanding the I-thou relationship as a unique illustration of the intimacy of primordial intersubjectivity, or inter-human relation. For Buber, '[t]he primary word *I-Thou* establishes the world of relation' in three 'spheres' or worlds: the natural, the human, and the spiritual (Buber 1958/2013: 5). It is the human sphere or the human world (*Mitwelt*) in which the relation established from the second-person perspective is *inter*personal or between two human individuals. It is a unique communicative relation in that the 'primary word *I-Thou* can be spoken only with the whole being … I become through my relation to the *Thou*; as I become *I*, I say *Thou*' (Buber 1958/2013*:* 9, emphasis original). In the vein of highlighting the 'betweenness' of human relationality as essential to human existence, Buber concludes: 'All real living is meeting' (Buber 1958/2013*:* 9). This shows that there is no detached first-personal self-awareness of a solipsistic nature. Rather, there

is a second-person perspective implicit in first-personal self-awareness: that is, I become *I* only insofar as I address the other as *you* and imagine the other addressing *me* in the same way, from the second-person perspective.

Stressing the reciprocity of the I-thou relation, Buber states: 'Relation is mutual. My *Thou* affects me, as I affect it' (Buber 1958/2013: 12). And the essential I-thou phenomenon in the human world is accorded the status of a fundamental ontological truth: 'In the beginning is relation' (13). How does Buber's fundamental ontological truth, secured from the second-person perspective, bear on social ontology? Does it go against individualism in the sense that, in Buber's view, the individual subject, the I, is not ontologically self-sufficient, stand-alone existent, but is always inextricably intertwined in the essential I-thou complex, of which the thou is another individual subject similarly ensconced in the same complex? Is this inevitably an example of thin sociality, of impoverished we-identity, of two singular subjects who are constitutively lodged in a communicative entwinement that defines second-personal intersubjective engagement? And, more importantly, is this dyadic *I-you* relational framework of interpersonal self-awareness explanatorily inadequate for a theory of thick sociality, or the multitudinous *I-we* relation? Is the I-thou relation engendered by the second-person perspective of consciousness, which is theoretically unsuited to conceptualising the nature of the multitudinous we, or the first person plural? Alternatively, might this kind of interrogation actually rest on a misapprehension of Buber's full contribution towards a theory of we-sociality?

Maurice Friedman has warned us against such a misapprehension and misreading of Buber's thoughts and referred to the availability of the concept of an 'essential *we*' in Buber's later works (Buber 1965; 2002). This concept, argues Friedman, serves as an important corrective to the mistaken notion that Buber's thought is limited to the relation between one person and another and does not encompass the larger realities of community living (Friedman 1965: 39). The misapprehension is to think that Buber's emphasis on the dyadic 'I-thou' relationship in second-person singular terms restricts him to a narrow philosophy of dialogue, with no scope for talking about the first-person plural understood in terms of the relation of the singular subject with multiple other persons in larger

communal realities. Referring to one's relation to a multiplicity of others, or experiencing robust we-sociality, Buber elucidates:

> By *We* I mean a community of several independent persons, who have reached a self and self-responsibility, the community resting on the basis of this self and self-responsibility, and being made possible by them. The special character of the *We* is shown in the essential relation existing, or arising temporarily, between its members; that is, in the holding sway within the *We* of an ontic directness which is the decisive presupposition of the *I-Thou* relation.
>
> (Buber 2002: 208)

We-ness is defined in terms of an 'essential relation' that exists between its members, and this essentiality is spelt out in terms of an ontically deep experience of 'directness' the members of a we-group feel amongst themselves. In other words, the we-subjective 'oneness' implied in the experience of 'one of us' that members feel towards one another in a multitudinous group is, after all, a sense of *being* one-of-us, or being a co-personal subject. Thus, a personal subject, defined by first-personal *singularity*, changes its ontic stance towards first-personal *plurality*, whereby it evolves into a co-personal subject. In other words, the occurrence of co-personal subjectivity—or we-subjectivity—through transformed ontic stance is what, for the I, it means to experience a sense of *communion* with others, who are all members of the we-community. Co-personal we-subjectivity is, in fact, a phenomenological space in which individual I-subjects feel as though they have unmediated ontic access to one another's being or self in the we-group. Each member's singular self-awareness—the awareness of being an I-subject—is overwhelmed by the collective consciousness of being a we-subject.

This idea of we-subject is analogous to what Hans Bernard Schmid calls the 'plural self' that consists in 'plural self-awareness'. Schmid argues that

> plural pre-reflective self-awareness constitutes a plural self and a plural perspective (we-perspective) in the same way in which an individual mind, and a first-person singular-perspective, is constituted by an individual 'sense of self', or … by individual pre-reflective self-awareness.
>
> (Schmid 2014: 12)

There is a pre-reflective sense of *us* anchored to a sort of mental integration in which the participant members of a we-group are aware of their attitudes as *ours*, collectively. The argument claims that self-awareness can also be first-personal plural, and the 'sense of *us*' represented in plural self-awareness is really constitutive of integrated collective attitudes similar to the way singular self-awareness is constitutive of an individual mind (Schmid 2014: 17).

If pre-reflective plural self-awareness is a basic fact just like singular self-awareness, then there is no question of the *formation* of we-sociality through any kind of relational process involving singular subjects. Even Buber's I-thou relation would, in this view, be an instance of plural self-awareness insofar as it represents minimal we-identity or the sense of being members of the dyadic 'you and I' group. At the same time, Schmid stresses that the 'we' must be understood as minds-in-relation rather than as some sort of undifferentiated unity. Since the we involves plurality, it would be absurd to conceive of it as some kind of 'larger-scale I' (Schmid 2009: 156). So, in this interpretation of we, there is no question of *bridging the gap* between the first-person singular and the first-person plural. The gap is not there because experiencing we-identity—having the sense of *us*ness—involves having plural self-awareness in complete absence of singular self-awareness.

If the 'we' is admittedly a plurality of *minds-in-relation*, then it must be a specific kind of social formation with an internal complexity that explains the specificity of this formation. Also, even if the participant members have plural self-awareness of shared attitudes and interests, individual participants have their self-understanding of interpersonal connectivity by way of experiencing oneself as *one of us*—as an organic part of a co-

subjective integration of multiple first-person singular perspectives. In some sense, the *I-we* nexus is a part-whole relationship: I experience myself as part of the 'we' I belong to, and am therefore self-conscious of my internality to the we-group. Describing this self-conscious realisation of we-identity as a matter of 'experiential anchoring' of the I to the we-community, Zahavi claims: 'To be part of a we, is to be with and relate to other prospective members in a distinct way; one involving participation and identification…. [Y]ou cannot be a member of a we without somehow affirming or endorsing that membership experientially. This is what makes the we a *first-person* plural' (Zahavi 2021: 13).

It is at this juncture that the complex structure of I-we relationality can be analytically discerned. As a part of the I-we nexus, I experientially affirm *my* being-with-others rather than bypassing this first-person *singular* self-awareness of group membership. At the same time, my self-awareness of being-with-others is obviously *inter*personal—hence a case of *plural* self-awareness. The I-we relational complex is actually an overlapping and intersecting space of both singular and plural self-awareness. We-sociality is not wholly explained merely by reference to its basis in plural self-awareness, nor does it bypass singular self-awareness as entirely extraneous to its formation. Rather, the two streams of self-awareness run together, with singular self-awareness more like a confluent undercurrent of plural self-awareness.

4 I, You, and Becoming We

If pre-reflective singular self-awareness is structurally embedded in pre-reflective plural self-awareness in the formation of we-sociality, then there must be some account of how the first-person singular perspective connects itself with the first-person plural perspective. I find one such account in Buber, especially when he says: 'Only men who are capable of truly saying *Thou* to one another can truly say *We* with one another' (Buber 2002: 208). This means that the natural ability and spontaneous propensity to adopt the *second-personal* stance towards the other foregrounds the I-we relation or membership in larger we-groups. Buber considers the I-thou relation as an 'essential relation' to assert its primacy in person-to-person relation.

In saying that '[t]he *We* includes the Thou potentially' (Buber 2002: 208), Buber stresses the point that the unmediated ontic intimacy that characterises the I-thou relation in its dyadic constellation is also involved in the multitudinous I-we relation. The second-personal stance of subjectivity is ingrained in the subjectivity of the first-personal plural stance because the microscopic ontic intimacy of the I with the thou is now macroscopically extended by the I to bring many other subjects within its phenomenological fold. In other words, the same ontic openness with which the I relates to another subject in the second-person singular engagement is enhanced when the I enters into a first-person plural engagement. In essence, the I-thou relation is a condition of possibility for the I-we relation. It is the mediation of the second-person perspective or engagement that foreshadows the subject's capacity for first-person plural engagement. It is in this sense that one can arguably consider the formation of we-identity to be a logical continuation and a phenomenological expansion of the essential I-thou relation.

If the I-we relation of sociality is logically continuous with the I-thou relation, and the latter form of sociality is a phenomenological expansion of the former, then the way I relate myself to a plurality of others in a we-group has strong qualitative affinity to the way I relate to a singular other. In that case, however, I see the members of the We, not as more-or-less accurate representatives or embodiments of the in-group prototype, but rather as individual persons in their particularity, much like the particularity of a You in the I-thou outlook. Even though the commonality of a group is defined by the accentuation of similarities or the prototypical characteristics of the members of the group, a genuine experience of we-identity—or identifying with the members of *us*—involves viewing the fellow members in their distinctive individuality as 'concrete others' who each can be addressed as a Buberian Thou.

There may be an immediate reaction against this formulation of the I-we relationship on the basis of the second-person standpoint, or through the mediation of that standpoint. The second-person singular communicative events make it evident to us that the intentionality of the you-stance is characterised by a highly personal focus on the concrete individuality of the second person. Therefore, it would seem that this second-personal communicative focus on individuality, tinged with a personal experiential

quality, is a feasible standpoint towards a multitudinous we-group only when the group is relatively small, such as a family or a circle of close friends.

Given these characteristics of the intentionality and phenomenology of second-personal intersubjectivity, the formulation of the sociality of the first-person plural stance through the mediation of the second-person standpoint looks infeasible when it comes to explaining membership in larger-sized we-groups, such as a nation or a global community. My first-personal plural stance towards millions of my fellow citizens of India would seem to be more grounded in a kind of impersonal, although emotional, affiliation to the socio-political identity of Indians in general. Beyond the scope of the fellow Indians that I personally know, and those that I know about, my we-stance would lack any focus on the concrete individuality of the other that I reckon as one of *us*. Indeed, there appears to be no room for viewing the enormous multitude of persons in their concrete particularities and personal idiosyncrasies.

Benedict Anderson has the term ‘imagined communities’ as a descriptor of a large social group—specifically nations. Anderson clarifies that a nation ‘is imagined because the members of even the smallest nation will never know most of their fellow-members, meet them, or even hear them, yet in the minds of each lives the image of their communion’ (Anderson 2006: 6). If I, as a member of such a nation, am never going to meet millions of my fellow national members, I have no opportunity for second-personal engagement with any of them. I would seem to have only an associative relation of co-existence with these others in millions as citizens of the same nation. Instead of nationalist affinity born of unmediated ontic intimacy, the envisaged scenario would seem to delineate the picture of a society of ontic *indifference*, severely lacking in the affective charge of we-identity.

A nation in such a condition runs the risk of disintegration. It may even be on the brink of collapse. And yet, a large nation may sustain itself despite the fact that millions of its fellow members remain ignorant of one another. When such a nation continues and even thrives into progress despite wide interpersonal ignorance, there is something that prevents ontic indifference from prevailing in the national scenario. This can happen largely because, even though one only imagines the existence of these millions of others,

one imagines them in terms of their existence as concrete individuals with particular problems and goals of life. In this ‘nationalist imagination’, the image of an unknown fellow national is not made up of an abstract prototype or an in-group representative figure. Rather, it is the image of a concrete person with whom one relates with openness, in the spirit of ontic intimacy, as one relates with a Thou. Only then does a fellow national member enter into the genuine circle of ‘we’.

Imagined communities can still be held together in the spirit of true we-identity, provided that the participant members imaginatively reach out to their multitudinous fellow members in their concrete particularities, and with the openness of ontic intimacy that characterises the I-thou relation. Thus, the second-person perspective is a structurally mediating link between the first-person singular and the first-person plural. This structural mediation is palpably true in the case of the formation of smaller we-groups, but seemingly absent in the formation of large-scale we-groups. The crux of the matter of we-sociality is that the I cannot recognise a large group of others as *experientially* related to itself, whereas the I’s felt recognition of the we-identity of the group as *us* is precisely what makes we-sociality a *first-personal* phenomenon. The possibility of this first-personal, experiential we-identity rests on the mediation of the second-personal I-thou experiential intimacy. This microscopic intersubjectivity of a dyadic constellation forms the foundation of the I’s further phenomenological expansion of first-person plural subjectivity to create the macroscopic nexus of robust we-sociality.

References

Anderson, Benedict. 2006. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, Revised Edition. London & New York: Verso.⏎

Buber, Martin. 1958/2013. *I am Thou*, trans. Ronald Gregor-Smith. London: Bloomsbury.⏎

Buber, Martin. 1965. *The Knowledge of Man: Selected Essays*. New York: Harper & Row.⏎
Buber, Martin. 2002. *Between Man and Man*, Second Edition, trans. Ronald Gregor-Smith. London & New York: Routledge.⏎
Carminada, Emanuele. 2016. 'Husserl on Groupings: Social Ontology and the Phenomenology of We-Intentionality', in Thomas Tzanto & Dermot Moran (eds.), *Phenomenology of Sociality: Discovering the "We"*, pp. 281–295. New York & London: Routledge.⏎
Friedman, Maurice. 1965. 'Introductory Essay', in Buber, M (ed.), *The Knowledge of Man: Selected Essays*, pp. 11–58. New York: Harper and Row.⏎
Gilbert, Margaret. 1989. *On Social Facts*. London: Routledge & Kegan Paul.
Gilbert, Margaret. 2003. 'The Structure of the Social Atom', in Frederick Schmitt (ed.), *Socializing Metaphysics*, pp. 39–64. Oxford: Rowman & Littlefield.
Gilbert, Margaret. 2007. 'Searle and Collective Intentions', in Savas Tsohatzidis (ed.), *Intentional Acts and Institutional Facts: Essays on John Searle's Social Ontology*, pp. 31–48. Dordrecht: Springer.
Heidegger, Martin. 1927 [1962]. *Being and Time*, trans. J. Macquarrie and E. Robinson. Oxford: Basil Blackwell.⏎
Heidegger, Martin. 1982. *The Basic Problems of Phenomenology*, trans. A. Hofstadter. Bloomington: Indiana University Press.⏎

Husserl, Edmund. 1989. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second Book. Studies in the Phenomenology of Constitution*, trans. R. Rojcewicz & A. Schuwer. Dordrecht: Kluwer Academic Publishers.⏎

Schmid, Hans Bernard. 2009. *Plural Action. Essays in Philosophy and Social Science*. Dordrecht: Springer.⏎

Schmid, Hans Bernard. 2014. 'Plural Self-awareness', *Phenomenology and Cognitive Science*, 13: 7–24.⏎

Searle, John. 1990. 'Collective Intentions and Actions', in Philip Cohen, Jerry Morgan, and Martha Pollack (eds.), *Intentions in Communication*, pp. 401–415. Cambridge: MIT Press.

Searle, John. 1995. *The Construction of Social Reality*. New York: The Free Press.

Searle, John. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.

Tollefsen, Deborah. 2004. 'Collective Intentionality', Internet Encyclopedia of Philosophy. Accessed on 15 January, 2026.

Zahavi, Dan. 2019. 'The Second-Person Engagement, Self-Alienation, and Group-Identification', *Topoi*, 38: 251–260. https://doi.org/10.1007/s11245-016-9444-6.⏎

Zahavi, Dan. 2020. 'We and I', *The Philosopher*, 108 (4): 19–24.⏎

Zahavi, Dan. 2021. 'We in Me or Me in We? Collective Intentionality and Selfhood', *Journal of Social Ontology*, 7(1): 1–20.⏎

OceanofPDF.com

13 First-Person Perspective and Collective Intentionality

A Relook into the Relationship between I and We

Manidipa Sen

DOI: 10.4324/9781003612827-14

The main aim of the chapter is to argue that without awareness of ourselves as the 'I' the very idea of 'We', that is, our awareness of ourselves as part of a larger group/collective cannot get off the ground. Hence prioritising the 'We' over the 'I' even within a group/collective may raise certain difficulties. The chapter tries to address this issue by considering the very notion of self-knowledge and, more significantly, the idea of first-person perspective, connected with the notion of self-knowledge. In doing so, the chapter tries to connect up some of the recent literature on self-knowledge with that of social ontology.

There is an overarching commonsense view regarding the inherent asymmetry between self-ascriptions of intentional mental states and ascriptions of intentional states to others due to the kind of authority we have over our self-ascriptions. The seat of this asymmetry lies in our having privileged access to our own mental states. But at the heart of these two interconnected ideas of asymmetry and privileged access lies the idea of first-person perspective/stance that we have *vis-à-vis* our mental life, a perspective or stance distinct from that of the stance that is widely known as the third-person or the objective stance we take while knowing others' mental states and the external world. Furthermore, the idea of first-person perspective and I-intentions go hand in hand and are essential for self-

knowledge. Having briefly set the grounds for talking about I-intentions and their significance, the chapter tries to articulate the role of self-knowledge and I-intentions in collective agency and We-intentions. The more important question that we would like to address here is whether primacy of the ‘We’ and We-intention undermine the very notion of ‘I’ and I-intention and thus the first-person knowledge of our own intentional mental states.

The chapter is divided into five sections. After having brought out the significance of I-intentions in the first section, we try to concentrate on the ‘priority debate’ between I and the We in the context of a collective in the second section. Following the articulation of what is meant by We-intentions, the third section concentrates on explaining the ‘Community first’ view, which is at the heart of We-intentions. In the fourth and fifth sections we critically evaluate the ‘Community first’ view and the priority of We-intentions over I-intentions to make a case that We-intentions lose their meaning without I-intentions and therefore, whether we experience ourselves as being part of a group or not, I-intention remains always the primary intention.

1 First-Person Perspective and the Significance of I-Intentions

In order to establish a connection between our experiencing ourselves as an individual in an I-mode and experiencing ourselves as part of a collective or community in a We-mode, we need to understand what constitutes the I-mode of experiencing. For this, we will begin by articulating the very nature of self-knowledge, that is, knowledge of ourselves in the I-mode, involving I-intentions.

Self-knowledge, that is knowledge of our own mental states, is considered necessary for our actions, and thus for us to be agents. This kind of knowledge has also been considered distinct from our knowledge of the external world and knowledge of other human beings (minds). Thus, the problem of self-knowledge is very often connected with the question of knowing our own minds, that is, knowing who we are in terms of knowing the beliefs, desires, wants, hopes, aspirations, etc. that we have, in general

knowing our own thoughts and feelings. Hence, the problem of self-knowledge can be rephrased in terms of the following two sets of questions:

1. Why is such a kind of knowledge important and significant for us? What role does it play in knowing ourselves as being distinct from knowing the mental states of others?
2. Why is such a kind of knowledge regarded as special and in some ways distinct compared to other forms of knowledge? Why do we think that knowledge that I am happy or depressed is distinct from my knowledge that the sky is blue or that the grass is green?

In answering the above two questions, one can say that questions concerning knowledge of one's own mental states are important not only because of their relevance for epistemology, but also because of their importance for human agency, freedom, and moral responsibility. Unless I am aware of my wants, desires, hopes, aspirations in an authoritative manner in a first-personal mode, my actions will lose their significance, and also lose, in certain contexts, their moral worth. More importantly, I cannot be held responsible for my actions unless I am aware of the beliefs, wants, wishes, etc. on the basis of which I act. However, it is not always clear how these two sets of issues, one mainly epistemological and the other mainly ethical, are connected with each other.[1]

Self-knowledge associated with this kind of commitment involves the following features that have been widely discussed: (1) The idea of asymmetry, (2) The idea of privileged access, and (3) The idea of first-person perspective. The overarching commonsense view is that there is an inherent asymmetry between self-ascriptions of intentional mental states and ascriptions of intentional states to others due to the kind of authority we have over our self-ascriptions. The seat of this authority lies, for the cognitivists, in the way we access our own mental states. Thus, the idea of privileged access, in terms of either some form of inner observation of our own mental states or some kind of agential privilege, emerges in the philosophical literature. But at the heart of these two interconnected ideas of asymmetry and privileged access is the idea of first-person perspective/stance that we have vis-à-vis our own mental life, a perspective or stance distinct from that of the stance that is widely known as the third-

person or the objective stance we can take while knowing the mental states of other human beings. One of the most important features of first-person perspective lies in taking my mental states *as mine*, the idea of 'mineness' that accompanies each and every piece of our thinking. It is only because we take our beliefs, desires, etc. to be our own that we are able to appeal to them in order to explain our agential commitments. Further, it is this first-person perspective that is preserved in our avowals (that is, first-person present tense self-ascriptions) like, 'I believe that drinking water is good for health', 'I intend to join piano classes', 'I want to help those in distress', and so on, that are articulations of our I-intentions. Hence, the idea of first-person perspective, their linguistic articulations as avowals and I-intentions go hand in hand and are essential for self-knowledge—in whichever mode we may take it to be.

Having briefly set the grounds for talking about I-intentions, we may try to articulate the role of self-knowledge and I-intentions in collective agency and We-intentions. The more important question that we would like to address here is whether primacy of the 'We' and We-intention undermine the very notion of 'I' and I-intention and thus the first-person knowledge of our own intentional mental states.

2 Role of the 'I' and the 'We' in the Priority Debate

We may cash out the problem mentioned in the previous section as follows: If we are to account for our experience of collective intentionality as preserved and maintained in and through fulfillment of We-intentions we need to address the nature and status of the We. But in order for us to be clear about the We, we need to be clear about the relationship between the I and the We. Understanding this relationship becomes even more important once we recognise that We and therefore We-intentions are not just an addition of particular I-s and I-intentions; and therefore, We-intentions are *sui generis.*

In understanding this relationship between the I and the We, there could be three possibilities available for us: (1) I is ontologically prior to We, and thus our first-person knowledge is prior to our knowledge as We, (2) We is

ontologically prior to I and thus our knowledge as We is prior to first-person knowledge, and (3) We and I are ontologically on a par and hence, We-awareness and I-awareness, as also corresponding We-intentionality and I-intentionality, co-emerge. But in order for us to choose between these three alternative positions, and decide whether We is actually prior to I—both ontologically as well as epistemologically-we need to be clear about the nature of the self itself, and the idea of first-person perspective that accompanies any notion of the self. Thus, the initial problematic that I articulated in the very beginning of the chapter can be paraphrased in terms of the following questions: Can we at all make a case for such a primacy of We-awareness? If we can, does that involve giving up on the idea of 'mineness' that accompanies our I-thoughts altogether? What, if at all anything, happens in such a loss? These are therefore the moot questions in deciphering the role of the I in taking ourselves to be part of a We community.

In order for us to understand in what sense We is prior to I, and thus We-intentionality is prior to I-intentionality, and therefore, an irreducible ontological category, we need to understand the ontological status of We and how we conceive of ourselves as We. This requires a further understanding of the very idea of having the 'same' intentions, or having the 'same' mental states, as part of a We-community.

Lynn Rudder Baker, in her paper 'Human Persons as Social Entities' (Baker 2015) explains what it is for an entity to be an independent and irreducible member of ontology. An entity is considered 'irreducible and ineliminable (at time t) if and only if it is not reducible and not eliminable (at time t)' (Baker 2015: 77). An entity is reducible if and only if it can be explained completely; that is, explained away in terms of something else. It is eliminable if it can be shown not to exist. According to Baker, Social Ontology is part of this general concept of ontology, comprising irreducible and ineliminable entities of a particular kind. Hence, according to Baker,

> Social ontology, as I construe it, is part of ontology simpliciter. As part of ontology, social ontology should include all social entities, social kinds and social properties that are irreducible and ineliminable. A social property is one

> for which social and linguistic communities are necessary for its instantiation. A community is one whose members bear significant intentional relations to one another. …A community is a group of persons with a measure of cohesion, with common intentional properties or relations such as shared interests, or values or language.
>
> (Baker 2015: 78)

This way of characterising social ontology, though rather vague, brings out certain significant features of a We-community. In the first place, not any group of people constitutes a social group, bound up, as it were, with some kind of We-awareness as an irreducible and real awareness. Secondly, it is constituted of members that not only have cohesion, but a kind of cohesion predicated upon having shared intentional mental states. Furthermore, this sharing of intentions cannot be reduced to having individual intentions, such that having individual intentions is in a way extraneous to having shared intentions. If that is the case, then the question still remains, how do individuals come together to form a social group that is, how does a cohesive group of individuals become a group that can be referred to as a We-community?

The reason why these questions are pivotal while trying to understand the nature of the We-community with shared intention and its relationship with the I lies in the following fact: Human beings, on the one hand, experience themselves as individuals, with their unique point of view on the world and other human beings, with a life of the mind and the body that is exclusive to the individual. They experience themselves as an individual with all its singularity. On the other hand, they also experience themselves as a social being, someone who is part of a family, community, or group, with a life of the mind and body necessarily socially connected to others, far and near. Furthermore, as individuals, one can forgo individual identity in favour of a larger social identity. We also experience ourselves as endorsing and prioritising our individual identity by giving up our social identity. Given that our experience of ourselves vacillates between these two perspectives—the individual and the social—it seems next to

impossible to privilege one identity over the other. On the other hand, if we do not privilege any one of the two perspectives, there is always a possibility of collapsing one to the other. Or one may say that this vacillation in our experience of ourselves as both individual and social has to be further substantiated by privileging one over the other precisely because we need to explain how our We-identity emerges if it is not just a summation of I-identities. And the best way to do so is by arguing that We is prior to I at an ontological level.

3 The 'Community First' View

Philosophers who are in favour of the 'community first' view—that is, a view arguing in favour of our social identity to be primary over our individual identity—can argue for their position in different ways. Dan Zahavi in his paper 'We in Me? Or Me in We? Collective Intentionality and Self-hood' (2021) explains different ways in which the primacy of We over I has been cashed out.

1. In the first place one can argue that, contrary to what philosophers have traditionally thought, our experience of ourselves as part of a community arises before we actually develop an individual or an I perspective on the world.
2. Our experience of ourselves as an individual is predicated upon the possession of a first-person concept, which, in turn, 'requires the capacity to conceive of oneself as oneself and the linguistic ability to use first-person pronoun to refer to oneself' (Zahavi 2021: 5) which further requires our being part of a linguistic community. Hence, unless and until we are already located in a defined linguistic community, we cannot even start conceiving of ourselves in the 'I' mode.
3. Thirdly, it can be argued that 'the community does not merely condition *what* we experience, but also *that* we experience. On this account, subjective experiences are social constructs' (Zahavi 2021), without having any independent epistemic or ontological basis. Our exclusive experiences make no sense precisely because

> they arise out of relations in the world. Hence the very fact that we experience is made possible because of our being part of a community.

One can see these three arguments as supplementing each other in the establishment of the ‘community first’ view. Hence Zahavi elaborates the community first view in the following manner. According to him, the community first view is based on the recognition that self always exists within a ‘normative space’ (Zahavi 2021: 6). A normative space is a space of social relationships, which further is the seat of values, goals, desires, and preferences. My development as an individual presupposes my being in this social space…the picture that ‘each of us develops our own preferences… in splendid isolation…is nothing but a fantasy’ (Zahavi 2021). It is a mere illusion to conceive of ourselves to be insulated individuals whose goals and preferences—that is, those that define us as a singular self—can be self-generating. That which is of significance and meaningful for me is shaped by the community of which I am a part. But this is not enough. We can still consider the I to be prior to the we within this space unless we further acknowledge, ‘my very identity is not something ready-made, something fixed by nature that simply awaits discovery. Rather, it is by forging an identity that I become a self’ (Zahavi 2021). Hence self or the I is not something given but something constructed in and through our being part of a community. Thus, according to these philosophers,

> to think that one can get to the core of human selfhood by abstracting away from the social context is consequently to fundamentally mischaracterise the relation between the I and the we. Rather than being an antecedently individuated self, rather than being merely contingently embedded in a community, my identity as an individual has a communal origin.
>
> (Zahavi 2021)

Since normative considerations are essential to who I am and since normativity is socially constituted, who I am is also socially constituted. Therefore, individual identity is parasitic upon community identity.

Another 'community first' view that Zahavi considers is proposed by Lynn Rudder Baker (2015). The idea that human self as person is socially constituted is argued for by distinguishing what she terms as 'rudimentary' from 'robust' first-person perspective (2015: 79). According to Baker, first-person perspective is a dispositional property that members belonging to the general kind 'person' essentially have. It is a dispositional property with two stages: the rudimentary and robust. Baker goes on to say that first-person perspective at a rudimentary stage is a 'non-conceptual capacity for intentional behavior that requires consciousness and intentionality' (2015). It is a capacity that human beings share with other non-human animals. However, as she further clarifies, 'For human persons, in contrast to non-human animals, the rudimentary first-person perspective comes with a piggybacked second-order capacity for a robust first-person perspective…' (2015) How does Baker characterise this second-order first-person perspective?

1. Firstly, it is unique to human beings.
2. Secondly, it is an ability to conceive of oneself as oneself.
3. Thirdly, it is an ability to distinguish self from non-self.
4. Fourthly, it helps in acquiring self-consciousness.
5. Finally, it is an ability that helps us in conceiving ourselves as agents responsible for our actions.

She further explains how we acquire this robust first-person perspective thus, 'One acquires robust first-person perspective as one acquires a natural language. To conceive of oneself as oneself in a first-person way one needs a self-concept…I take concepts to be inextricably linked to language' (2015: 79–80). Acquiring any concept, and therefore acquiring a self-concept, according to Baker, is predicated upon acquisition of language. Hence the robust sense of first-person can arise only when someone is part of a linguistic community. Hence the dispositional property of first-person perspective belongs to social ontology.

From the above discussion, it becomes evident that there is an inextricable relation between I and We. It is also evident that, as a result of this, there is an inextricable relationship between I-intentionality and We-intentionality. But the question that still remains to be asked is whether We-community and, therefore, We-intentionality is necessarily prior to the I or I-intentionality.

4 The ‘Community First’ View and the Priority Debate Revisited

Let us note one important point here. The ‘community first’ view has to conceive of the self in a radically different manner so that it can argue that the ontology of the individual self itself has to be understood as being part of social ontology. It is not enough to hold that the subject or the individual consciousness or the self arises in a social context. An individual’s living in a community does not imply that community is prior to the individual. That is why those who are ardent defenders of the ‘community first’ view find difficulties in many of the social constructivist views suggesting that they suffer from an ‘individualist’ bias; that is, they are not able to subsume our experience of ourselves as an I in the context of our being part of a group or a we community.

Even those who are in favour of an autobiographical, narrative or an extended self cannot completely give up the idea of the ‘mineness’ of our experience. Thus, even when they accept that some dimensions of the self are clearly social (like the robust first-person perspective, as suggested by Baker), constituted, as it were, through our being in a community, they do not reject the idea that ‘there are other, arguably more fundamental dimensions that are present from very early on and which are linked to our embodiment and experiential life’ (Zahavi 2021: 9). Minimal form of selfhood is part of our experiential life, part of our subjectivity that can’t be subsumed under any social existence. Given this influential view about the nature of selfhood, those who adhere to the community-first view have to argue for a position stronger than what someone like Baker would have argued for. According to Zahavi, these philosophers have to contend that ‘phenomenal consciousness is constitutively dependent upon social

interaction. Not merely when it comes to its specific content, but as regards its very being' (2021: 12).

It is evident that having shared content or shared experiences does not necessarily imply a 'community first' view. We can be very much individuals in our own right even when we share our experiences and content with others. One can understand this approach in the following way. We may experience pain, grief, discrimination, etc., as part of a group without experiencing them in exactly the same way. The way one experiences discrimination by being part of a community, though similar, can never be the same. This can happen only if we allow for the possibility that the mode of experiencing pain, grief, etc. is different in individuals belonging to the same group. Therefore, the mode of experiencing is bound up with our own point of view, 'how it is for me to experience' the way I do. The same argument can be given to show the difference between individual beliefs as part of a community of believers. Individuals may be part of a community who believe that God exists without believing so in exactly the same manner. Thus, sharing of beliefs and their content as part of a community does not imply *sameness* of beliefs and their content. If it were so we would have to give up the possibility of disagreement, dispute, revision, and the further possibility of dissociating ourselves from the group we belong to.

Therefore, those who maintain a 'community first' view and prioritise We-intentionality over I-intentionality cannot thus base their view just on individuals having a shared content. Furthermore, just as shared content does not necessarily imply a community first view, being part of a group does not guarantee primacy of We-intentionality. We can be a part of a group or a community of people without actually sharing any significant beliefs with any of the other members of the group or community. I can be part of a community of bald-headed men or red-headed women without having anything else to bring us together. As Zahavi points out, 'you could be a full-fledged member of certain groups, say, the group of people with blood type O, regardless of whether or not you know it or care about it' (2021).

Hence being part of any amorphous group or community is not enough to establish the 'community first' view. In order for the 'community first' view to work, it is also not enough to accept the fact that human beings are

socially connected. Primacy of the We-community and thus We-intentionality requires a stronger claim than mere sociality. As required by the 'community first' view, it is important that the I is constitutively dependent on the We, that is, constituted by the We. As Zahavi points out, 'one should not make the mistake of assuming that one can prove the primacy of communal We, simply by pointing to the importance of sociality'. According to him, sociality does not merely involve sharing; it also 'covers relations of hostility, indifference, instrumental interaction etc.' (2021: 15). Hence mere social formations do not imply communal We, and the idea of a Communal We is essential for the 'community first' view to work.

For it to be a genuine group or community, individuals have to endorse their membership to the group. Furthermore, in order for me to be part of a community such an endorsement from the first-person perspective, though necessary, may not be sufficient. Whether I can be counted as a true member of a community depends also on whether others in the community recognise me to be so. In order for such an endorsement to happen, one has to be related to other possible members in a unique manner. That relation is further predicated upon experiencing oneself together with others. This point is evident from the fact that 'we' is a first-person plural pronoun, which can be understood in contrast to the first-person singular pronoun 'I'. Hence there is a need for this We-community to have related but, at the same time, distinct members. As Zahavi explains, 'We is constituted by me and at least one co-subject. But for me to relate to other co-subject is precisely to relate to someone who not only has a perspective on his or her own world, but on me as well' (2021: 17). We-ness of a community cannot therefore erase the I-ness within members of a community. Neither can it erase the distinction between the first and third-person perspective.

Based on these observations we can reach the following conclusion. It seems hardly possible for there being an experientiality of We unless and until we have an experientiality of the I. It is hardly understandable that experiences are We-experiences before they are I-experience. It can only be an experience of I-with-others. As it is evident from the discussion, the primacy of the communal We does follow from our being socially grounded. To quote Zahavi once again, 'Whereas different forms of we (a family, a soccer team, a religious group etc.) are all quite particular social

formations, sociality is a much wider umbrella term that also covers relations of hostility, indifference, instrumental interactions etc.' (2021: 15). Given all these possibilities the 'community first' view seems to face difficulties.

Hence the primacy of the first-person perspective and 'mineness' of experience cannot be given up in favour of a we perspective and 'We-ness' of experience. Our first-person perspective gives me a grounding on which we experience ourselves in relational terms. Primacy of 'I'-intentions does not undermine the importance of 'we'-intentions in human lives. The first-person stance on ourselves may become more meaningful once it is understood in relational terms, though that does not support the 'community first' view on ourselves.

5 Reviewing the Ontological Priority of We

The 'community first view' has to commit itself to and also argue for the following points: (1) The We is ontologically prior to the I; and (2) Our knowledge and experience of ourselves as We is prior to and necessary for our knowledge and experience of ourselves as I. In order to make sense of these two points, we may look at the way the pronouns, like 'I' and 'We', are used in language.

A lot of discussion in contemporary literature on knowledge of our own mental states, and thus knowledge of who I am as an agent with specific intentions, stems from the asymmetry that philosophers of language and mind have observed between first-person self-ascriptive statements like 'I am happy' or 'I believe that drinking water is good for health' (known as avowals) and third-person ascriptions of mental states 'Manidipa is happy' or 'Manidipa believes that drinking water is good for health'. If we are to make any headway in trying to understand the complex relationship between the I and the We, that is, if we are to make a case for a truly 'community first view', we need to understand the use of 'I' and 'We' as first-person singular and first-person plural pronouns. One prominent reason why self-ascriptions involving the use of 'I' is distinct from third-person ascriptions involving the use of proper names like 'Manidipa' is that in referring to the subject in the first-person mode, the pronoun is immune

to error in two senses: (1) ‘I’ cannot misrefer; (2) ‘I’ cannot lack reference. As succinctly articulated by Strawson, ‘If someone has, and formulates to himself or herself, such a thought as ‘I am feeling terrible’, then his or her ‘I’ is guaranteed against two kinds of failures to which the uses of some other definite referring expressions are sometimes exposed: it is guaranteed against lack of reference, and it is guaranteed against mistaken or incorrect reference’ (Strawson 1994: 210).

This linguistic feature of ‘I’ has been used to argue for some special features of self-knowledge as opposed to knowledge of other minds. Some of these special features, as mentioned earlier, are Immediacy (i.e., no independently articulatable grounds for claiming to know one’s own mind are required), Authority (i.e. we know best what is going on in our own minds), and Transparency (i.e. the mental state that one has is one that the person knows as one’s own).

These are the features that distinguish self-knowledge from knowledge of other minds and knowledge of the external world, and hence the claim of asymmetry between self-knowledge and other forms of knowledge. As Crispin Wright and others point out in the introduction to their anthology *Knowing Our Own Minds* (2000), though there is controversy regarding questions like: How is it that our knowledge of our present mental states has the following features? or How are the conditions of immediacy, authority and transparency satisfied by my saying (avowing) that I am in pain at this point in time?, these minimal features of self-knowledge have been widely accepted by philosophers. This epistemic asymmetry has further been used to argue for a special ontological status for the first-person, the subject, the I. This is also the locus of one’s subjectivity, a special perspective in terms of which we know and relate to the world and other human beings.

If we are to privilege the ‘We’ over this notion of the ‘I’, both ontologically as well as epistemically, we may try to look at the use of first-person plural in our language. Use of ‘We’, in ascriptive statements like ‘We are happy’ or ‘We believe that drinking water is good for health’, is not immune to error in either of the two senses just mentioned, (1) it may very well misrefer to a collective other than the one intended, and (2) it may lack reference in case there is, in actuality, the sense that the collective is just an illusion. One can point out that since such mistakes are possible vis-à-vis We, as opposed to I, one can further argue our We-awareness does not have

the same status as I-awareness. Does it also indicate that We-awareness is not primary in the way it needs to be in order for us to make a case for the 'community first' view?

As mentioned earlier, those who take 'We' (being a collective) to be prior to 'I' (being an individual) would have to argue that the very awareness or knowledge of 'I' is socially constructed and hence dependent on awareness of We. Can a merely socially constructed awareness of 'I' and hence socially constructed I-consciousness fulfill the kind of authority that is demanded of the notion of 'I'? I think not. Let us see why. The immediacy, transparency and authority of our first-person self-knowledge suggest that I-awareness can't be a pure social construction based on our We-awareness, since knowledge of the I is so very distinct from the knowledge of We. Furthermore, avowals, that is first-person self-ascriptive statements, belong to a language game very different from that of statements of we-ascriptions. We can privilege the community-first view only at the cost of giving up on the idea of the primacy of our first-person awareness, which is impossible.

One of the arguments given by the community first view rests on the idea that the very awareness of the I, and thus the first-person perspective, is socially constituted because of the linguistic community we belong to. Our being a language user makes us necessarily part of a community, since our sense of 'I' and use of the pronoun 'I' arise only if we are part of the community of language users.

I would like to end this discussion by challenging this idea from the point of Davidson (2001).[2] For Davidson, to be a competent language user is to acquire the ability to interpret each other's speech. This requirement would be undeniably true for a community-first view as well. However, this kind of interpretation cannot get started if we are unable to take each other to be authoritative about our own mental states. The aim of the speaker is to make himself understood by the hearer. However, this process cannot even get started if the hearer does not take the speaker to know what she says and believes. The speaker need not, and cannot, interpret her own words to understand what they mean, while the hearer needs to do that all the time. This presumption is necessary in order for us to understand as well as interpret the speaker, and then ascribe the appropriate mental state to

him/her based on the clues provided by her. Therefore, there is no corresponding presumption that the interpreter knows what the speaker means and believes. Hence the utterer's knowledge about her own mental states and the interpreter's knowledge of the utterer's mental states is asymmetrical.

Taking clue from this idea, one can say that in a genuine We community, each member has to take the other as authoritative about what she believes, desires, wants, hates, loves, fears etc., based on the avowals (i.e., first-person self–ascription) uttered by them. In order for us to take each other authoritative about our own mental life in this manner, we have to take the first-person perspective on our mental life seriously. This, in turn, suggests that the I perspective, that is, the first-person perspective has to be present in order for us to forge a genuine We-community. Our use of language itself suggests that.

Notes

1. In a previous chapter I have tried to show the connection between the two domains through the works of Akeel Bilgrami, Richard Moran and Victoria McGeer. These philosophers, in very different ways, base the idea of self-knowledge in the context of agential commitment. See Sen (2023) in Ranjan Panda (ed.), *Self-knowledge and Moral Identity.*⏎
2. See particularly, 'First Person Authority' (pp. 3–15), 'Rational Animal' (pp. 95–106), 'Second Person' (pp. 107–123).⏎

References

Baker, Lynn Rudder. 2015. 'Human Persons as Social Entities', *Journal of Social Ontology*, 1(1): 77–87.⏎

Davidson, Donald. 2001. *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.⏎

Sen, Manidipa. 2023. ‘Self-knowledge and Agency: Revisiting the Commitment Model of Self-knowledge’, in Ranjan Panda (ed.), *Self-knowledge and Moral Identity*, pp. 199–211. New Delhi: Tulika Books.⏎

Strawson, P.F. 1994. ‘*First-Person and Others*’, in Quassim Cassam (ed.), *Self-Knowledge*, pp. 210–215. Oxford: Oxford University Press.⏎

Wright, Crispin, Barry Smith and Cynthia McDonald (eds.). 2000. *Knowing Our Own Minds*, ‘Introduction’, pp. 1–12. Oxford: Oxford University Press.⏎

Zahavi, Dan. 2021. ‘We in Me or Me in We? Collective Intentionality and Selfhood’, *Journal of Social Ontology*, 7(1): 1–20.⏎

OceanofPDF.com

Index

Note: Page numbers with ‘n’ refer to note numbers.

accountability gap 90, 92
accountability, forward-looking and backward-looking 90, 92, 102, 103
Adams, Nate: question of legitimacy 134
agency, individual agency, addressee’s 7, 242
agent perspective, agent-perspectival feature 216–218
aggregate needs 35, 37, 43
all things considered judgments 67–68, 71–72
Ambedkar, B. R. 193
Anderson, Benedict 238
Anderson, Elizabeth 187–188, 191–193, 196, 198n12, 198n16
anti-realism 17, 21–22
artificial persons, group agents as 65
assurance game 162
Ásta 2, 14
asymmetry (self-ascription *vs*. third-person ascription) 241, 243; epistemic 251
atomism 6
Austin, John L. 46, 47–48, 52, 55–56: on illocutionary *vs*. perlocutionary goals 47; on ordering 48–49
authority: self-knowledge 250; of ‘I’ 251

Bacharach, Michael 212, 215
Baker, Lynn Rudder 244: on social ontology 244; first-person perspective 247, 248
basic joint actions 208
being: I and we 226
Berger, Peter L. 163
best-reply reasoning 210, 213, 214, 216–218
bias: cognitive, implicit 185; confirmation 187; self-serving 192
Petersson Björn 127n24, 202, 209–210, 215–218

blame; appropriateness of 100; blameworthiness, quantum of blame error 65, 89, 92, 94–106, 110, 117, 120, 125n4, 153
brain-scientist example and social reality 3
Bratman, Michael 80: Bratman-lite 13, 202, 204, 208, 213, 218; Bratmanian co-operator 215–218
Buber, Martin: I-Thou relation 234, 237; and we-sociality 234
Buchanan, Allen: role of legitimacy 133; first-order coordination problem 133; criteria of legitimacy 141
business ethics 137
business-corporations 131–132, 138; and states 130

Carminada, Emanuele 226
characteristic goods, services 138
CMA thesis (Collective Moral Autonomy Thesis) 93
Coase, Ronald 13
coercion 99, 204–205
cognitivism 9, 19–22
coherence 184–185, 197n4
cohesion, social 244
collective activity, practices 9, 20, 24
collective agency 11, 130, 243
collective agents: business-corporations 130; organizations 130; and moral agents 143n1
collective demand 124
collective end 31–44
Collective End Theory (CET) 32–33
collective goods 31, 34, 35, 37, 42, 44
collective guilt 123
collective intentionality 163, 243: essence of 5, 8; and self-knowledge 14
collective reasoning 182–183, 188–191, 193, 196–197, 198n13
collective responsibility: possibility 8; backward-looking 91; non-distributive 89, 91, 95, 100–106, 106n1, 202
collectivity, idea of 5
Collins, Stephanie 130
commitment coordination 163; scope of 163; global 164; and epistemic practice 164
commitment, normative 70, 73, 75
common knowledge condition, we-intention and shared intention 207, 220n10
communication: communicative goal *vs*. illocutionary goal 49–50; as joint/plural action 53, 56–58; miscommunication 57–58
Community First view 245, 249; and priority of We 250
complete *vs*. explicit content 221n14
conferralism, conferred statuses 16, 23, 29

consciousness 13: first person perspective 226; plural subject of 227; and intentionality 227; I and We consciousness 228; collective 235
consequentialism, goals of 86–87; and team reasoning 177
consistency (rationality requirement) 67
Constituent Assembly, Indian 182–183, 193–197
constitutive agreement: Sellars 169; nature of 169; and coordination 170; unshared group attitudes 172; and deontological reasoning 175
constructionism, social 9, 18, 21n2, 27
content approach 203, 209, 213, 217–218
coordination 82n7, 112n3, 124, 151n2, 208–219
coordination games 161–162
corporate legitimacy 138, 141
corporate structure 130
corporations and states 140–141
corporations, as loci of responsibility 85
correlativity thesis 90–91, 103
costs: preventive; reparative 104, 106
coyly consenting 59, 61; Mark and Emily vignette 60–61

Davidson, Donald 251
deception 204
decision procedure 125n6
degrees of participation 151
democracy and state-firm analogy 143n2
democratic moral inquiry 188, 192, 198n12
deontology 161, Sellars's 178
derivative obligation 120
determinism, freedom and 77
dialogical space 229
distributivist interpretation; and non distributive accounts 90–91, 94, 109–110, 124
doxastic shared end condition 206–207
duties, reparative 105; welfare-related 106, 142

einhacken/*einhaken* (typo dispute) 51
elderly person drowning example 113
elections as collective actions 149
eliminativism 21, 77
epistemic access, limitations 4
epistemic objectivity 21
epistemic practice, function and structure 165
epistemic relativity 165
Error theory 19–20

first-person perspective: robust 246–247; rudimentary 247
first-person point of view: agency 69; normative construal of 71; phenomenological construal of 70
Fisher, Sarah A. 46, 59–61
Footballers' Problem 210–211
forward-looking obligation 110
Frankfurt, Harry 87
Freedom and Resentment (Strawson) 64
Frege style cases 206–207
Friedman, Maurice 234–235
functions 136
fuzzy boundaries 154

game theory 161–162, 209, 210
gap in the accounting-books94, 96–98
gender, as social kind 16–18, 149–152
Gilbert, Margaret 80, 130, 173, 225
grounding, metaphysical 20, 28
group agency; hybrid case of 7, 64–65, 83, 143n1
group agent approach 220n6
group agents: *vs*. natural persons 65; loci of responsibility 65
group identification 212
group intentions, unshared 176
group persons 65, 69, 74
group point of view 171
group/collective responsibility; non-distributive 90–91, 98

Habermas, Jürgen 53–54, 56, 58; theory of communicative action, claim to validity 58
Hart, H. L. A. 31
Haslanger, Sally 17–29
Heath, Joseph 137
Heidegger, Martin: Dasein 232; being-with and being-among 232
Held, Virginia 109, 110, 123; Kitty Genovese example 123
Herald of Free Enterprise 92–95, 97
herd immunity 166
Hi-Lo game 210–213, 215–216
Hobbes, Thomas 63
Hollywood Standoff example 94
How to Do Things with Words (Austin) 55–56
Hume, David 77
Husserl, Edmund 13, 48: communicative intertwinement 231
hybrid case of group/multiple personhood 82

I-consciousness 228
illegitimacy, corporate 143; and social license 143
illocution 46–47: as plural action 46, 52, 56–58, 60–61; speaker bias in classical theory 46, 55–56; illocutionary goal 47, 50; distinguished from communicative goal 49–50; distinguished from perlocutionary goal 47
illocutionary intention 46–58; as joint intention 56–58; speaker-intention *vs*. uptake-intention 54–56
illocutionary success 48, 50, 52, 56
implicit coordination/quasi-joint action 151
implicit/tacit intentions 208
incompatibilism 77
Incompatibility Thesis 202, 210
individual agency: condition of 64; and rational unity 69
individual rationality: overarching requirement of 72; specific requirements of 67
individualism and collectivism, tension 6
individualism, forms of 6
individualist interpretation; non-individualist interpretation 115, 125n10, 220n5, 186–187
individualist moral orientation 79
individualistic biases 2
inquiry, analytical 1
institutional agents 152
institutional rights 35
institutional roles 19, 30, 36–37
institutions and first-order coordination problem 133
instrumentalism 19
integrity, personal 79
Intention Condition 13
intentional social transformation 26
intentions: as reason-preserving 208; that we J 204, 205, 207; doing one's part 213
inter-institutional rights 41–44
interpersonal relations 73
intersubjectivism, intersubjectivity 23, 223–239
intra-personal relations 65, 73
I-Thou: relation 229–230; sociality 230; and intersubjectivity 234; dyadic relation 235; and I-We relation 237; *see also* Buber

Jensen, Michael 137
joint enterprises 34, 35
joint intentional action/activity 152, 201, 219n2
joint intentions and ontogeny 163
joint productive activity 30, 32, 38, 44
joint responsibility, and joint action 111–112, 117

Koons, Jeremy Randel 174; Labour 38–44, 131–137

Laitinen, Arto 29
legitimacy standards: subject of 134; non-minimal 141–142
legitimacy: crisis of 129; concept of 130; and territory 132; and metacoordination problem 133; general 135; of states 135–136, 138; and business corporations 133–134; functions 136; moral 140
Leviathan 7
Lewis, David 154, 155
liberty rights, obligations generated by 90
limited liability 132
List and Pettit 63: humans as natural persons 65; structure of group agents 81
Locke, John 38, 65–70: distinction between personal and animal identity 65; sameness of consciousness, Prince and Cobbler thought experiment 66
looping effect 26

Manhattan Project 84
manhole/flash-flood rescue case (three versions) 113–114
Manyism 154, 157
market failure theory 137
Martens and Roelofs 151
May, Larry; shared responsibility 109
McMahon, Christopher 188–192, 196
McPhail, Will; *Bedtime* cartoon 16, 25
Meckling, William 137
mereology 154
metacoordination problem 133
metaphysical reality 8
metaphysical reality 8
Miller, David: duty of rescue 118; remedial responsibility 126n17, 126n18
mind-body dualism 154
minimal realism: versions R1, R2 19, 22, 26
minimal sociality 230
mode approach 203, 213, 220n6
moral legitimacy 135–136
moral normativity 30, 32, 44
moral norms and team reasoning 169
moral progress: conception of 182–183; modeling of 183–186; and collectives 188–189, 192–193, 196–197
moral reasoning: game-theoretic approaches 161; and constitutive agreement 168; and unshared moral intention178
moral reasons, scepticism about 168

moral responsibility: genuine 111; object of 113
moral significance, of individual human beings 79
multiple personhood 82, 85
mutual recognition 63–64
mutualism 167

nationalist imagination 238
natural kind *vs*. social kind 17–21
Nietzschean instincts 56
non-agential groups, obligation bearers 89, 91–94
non-arbitrariness, obligation bearers; 91, 102
non-cooperative action 167–168
normative efficacy 30–44, 119
normative reduction 69, 71, 74
normative subjecthood 110
normative teleological account 30, 32
Nozick, Robert 38

objective stance 79
objectivist realism 16–29
ontology of the social 244
Oppenheimer, Robert J. 84
oppression; as collective act 25, 26, 149
organizations 130: realism 131; and agency 131; private 139
Ostrom, Eleanor 166–167
ought implies can 90, 112, 116, 125n9
overlapping individuals 154
own-action condition 203–204, 220n6

parallel action 204
Parfit, Derek: reductionist account of personal identity 66; teletransporter thought experiment 66
perlocution 47, 55
personal identity: reductionist account 66; and animal identity 67; animalist account 68; normative account 69, 74, 77, 79, 86
personal responsibility: forward-looking conception 87; retrospective conception of 87
persons: self-constituting nature of 75; smaller-than-human-size 75
perspective as indexical 226
physical proximity and duty of rescue 118–120, 127n23
Plato 63
plural action view 46, 52: Fondue example 57; handshake example 56; and joint intention 56–58

plural subject: we-subjectivity 225; existence of 227
Pogge, Thomas 91
political legitimacy 130
positive and negative rights 91
practice-dependence 16–29
practices: coordinating 162, non-coordinating and morality 165, 167; partially-coordinating 162, and reification 164
praise and blame: reactive attitudes 64, 78; practice of 87
predictive *vs*. deliberative stance 217
primary and secondary laws 31
priority debate (I *vs*. We) 243, 247
problem of the many/puzzles of multiplicity 154
productivity theory 137
protest as collective act 152–153
proxy agency 121
prudential reasons 168
putative collective 114, 126n13

quasi-representative action 110; representative action, character 110, 117–123

random collectives 11, 109–111, 125n8
rational unity, overall 67–69, 72, 84
rationalist moral theories 79
Rawls, John 176, 182, 184–186, 189–190, 197n4, 197n9, 198n14, 198n15
realism, success theory 16–29
reasoning transformation 211
reasons-responsive mechanisms 208, 214
reciprocity 30–45, 229–234
reflective equilibrium 12: criticisms of 183, 185–186, 191; individualistic interpretation of 186–187, 196, 197n8; narrow and wide 185, 189; structural elements of 183–184; variants of 184, 197n2
Reinach, Adolf 48–52, 61–62; on social acts 48; spear metaphor 49–51; ‘throwing to’ *vs*. ‘throwing at’ 50; typo dispute (*einhacken*/*einhaken*) 51
remedial responsibility 126n18
responsibility gap 90–92, 153
rights, basic 130
rights, needs-based 41–44; reciprocal 40–44, 206, 231
Rotblat, Joseph 84
Roth, Abraham Sesshu 202, 203, 213, 214
Rousseau, Jean-Jacques 63

sameness of consciousness 66, 85

Schmid, Hans Bernhard: plural self-awareness 235; sense of us 236; decision procedure 120, 124n2, 235
scientistic view 78, 82
Searle, John 7, 33, 163, 201, 203–206, 225; we-mode/I-mode 225
second-personal engagement 234
second-personal stance 237
self-ascription 250; and avowals 251
self-knowledge: problem of 242; and collective agency 243; and other minds 250; special features 250
Sellars, Wilfrid 12: we-intentions 127n26; Whooping Crane Society 169; and moral reasoning 171; teleology and Kant 176, 178; and utilitarianism 177; moral point of view 177; account of team reasoning 179
semantic reasons 168
shared deliberation 190, 207, 209, 212, 217
shared-token-intention approach 220n6
shareholders and business-corporation 131
silencing, of addressee 46
slavery, abolition of 182–183, 186, 192–193
social change 17, 25–27
social construction 17, 23, 24, 149, 251
social coordination, large-scale 164
social critique 16–29
social entities 1–14, 63, 81, 129, 134, 244
social facts, practice dependent 19–20, 28
social groups 5
social identity approach in social psychology 212
social institutions 41, 42, 44
social kinds 16–18, 22–23, 25–26
social moral epistemology 188, 192
social ontology 1, 2, 182–183, 188, 197, 197n10, 225
social phenomena: minimal forms of sociality 63; *vs*. non-social group agency 65
social reality: brain-scientist example 3; and reducibility 4; practice dependent 3, 6, 16, 22–25, 28, 225
Social Reflective Equilibrium (Social RE) 183, 186–191, 193, 197
social rules 30
social turn 2, 14n1
speaker bias 46, 55–56
speech act example: ordering 47–49; promising 59–60; telling 46–48, 52, 59; warning 47, 53–55, 58–59
speech act theory: classical view 46–49, 55–56; Habermas's version 53–54, 56, 58; *vs*. uptake act theory 55–56
stakeholders 132, 140, 142

state-firm analogy 143–144n2
Strawson, P. F: reactive attitudes 64; participant stance 78; non-reductive naturalism 79; freedom and resentment 8, 10, 64, 77–80, 87, 250
subject of intention/reasoning *vs*. the ontological subject 212
subjective internalization 233
subsistence rights; obligations arising from 89
suffrage, universal adult 182, 194–195

Tam, Agnes 193–196
team benefactor 210, 214, 216
team commitments 179
team preferences 210, 214
team reasoning 12: deontological 172; principled 167, 173; and morality 168, 171
teletransporter thought experiment *see* Parfit, Derek
theory of the firm 137
Threat Principle 99
token *vs*. type mind-dependence 16–29
Tollefsen, Deborah 228
Tomasello, Michael 163
tragedy of the commons 166
transaction cost theory 137
truth-aptness 19–25
Tuomela, Raimo: collective acceptance 172; ethos 174

unshared action 179
unshared commitments 179
untouchability, prohibition of 183
uptake intuitions, study of 46, 59–61
uptake, achievement 47, 54
uptake: as active *vs*. passive 52–53; addressee's agency, not speaker's 47, 50, 52–53; failure 49, 55, 59–61; as intended in speech (Section 3.1) 47–52; intention 54–56, 58
utility 189, 211

vernehmungsbedürftig/*Vernehmung* 48, 52

we-attitudes 180n6
we-community: formation of 245; and I 245; primacy of 249
we-consciousness: locus 227; emergentist interpretation 228; phenomenological unity 228; immanentist 226
we-intentions 7; sui generis 243
we-perspective 118, 120
we-reasoning 193–196

we-subjectivity, problem of 225
whistleblowing 143
Wiggins, David 35, 37
Wilson, David Sloan 163
Wittgenstein, Ludwig 14n5
women, in the Constituent Assembly 195–196
Wright, Crispin 251
Wringe-Korsgaard argument 89–92, 101, 104

Zahavi, Dan: interpersonal engagement 230; and Husserl 231; we membership 236; primacy of we over I 245; and Baker's community first view 246; minimal selfhood 248
Zuwerfen (cooperative throwing) 50

OceanofPDF.com