A Theory of Shared Agency

Michael E. Bratman
Stanford University

Draft – please do not quote without author's permission

Summer 2011
File: BratmanSharedAgencyFull.summer11.6.doc
9-6-11

This is an unfinished draft and is subject to revision. References and acknowledgments are incomplete. Comments are welcomed. Please do not quote.

Preface

I began my 1987 book, Intention, Plans, and Practical Reason, with a question:

> What happens to our conception of mind and rational agency when we take seriously future-directed intentions and plans and their roles as inputs into further practical reasoning?[1]

The answer that emerged, both in that book and in later developments, was that we are led to a way of thinking about our rational agency that departs in important ways from what was then commonly called the desire-belief model of our agency. And I came to call the resulting, overall model the planning theory of intention and our agency.

In the present book I turn to a closely related follow-up question:

> What happens to our understanding of small scale cases of acting together in ways that count as shared intentional and/or shared cooperative activity -- examples include singing duets, dancing together, having a conversation together, painting a house together -- when we take seriously the model of our individual agency provided by the planning theory?

And what I have come to believe, in reflecting on this question, is that we are thereby led to a rich and promising model of robust forms of small scale shared intentional agency. This planning model of our shared intentional agency aims to provide an illuminating reduction of such sociality to structures of interconnected planning agency. My main aim in this book is to lay out this planning model of our sociality with sufficient clarity so that we can see how it relates to other approaches, and so that we can try to reach an assessment of its general merits and usefulness for further research. A guiding idea is that this planning model helps us forge a plausible path between approaches that, in my judgment, are less theoretically attractive. These approaches include a simple reduction to strategic interaction, an appeal to a fundamentally new element of the mind, and an appeal to a fundamentally new form of normative inter-relation between persons. [more…]

Many of the ideas to be presented here were first developed in a series of six essays originally published between 1992 and 2006. These essays are: "Shared

[1] p. vii.

Cooperative Activity," "Shared Intention," "Shared Intention and Mutual Obligation," and "I Intend that We J" – all reprinted in my Faces of Intention (New York: Cambridge University Press, 1999); "Shared Valuing and Frameworks for Practical Reasoning," as reprinted in my Structures of Agency: Essays (New York: Oxford University Press, 2007); and "Dynamics of Sociality," Midwest Studies in Philosophy: Shared Intentions and Collective Responsibility XXX (2006): 1-15. Three more recent overview papers of mine – papers I draw on in various ways in this book -- are: "Shared Agency," in Chris Mantzavinos, ed. Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice (Cambridge: Cambridge University Press, 2009): 41-59; "Modest Sociality and the Distinctiveness of Intention," Philosophical Studies 144 (2009): 149-65; and "Agency, Time, and Sociality," Proceedings and Addresses of the American Philosophical Association 84:2 (2010): 7-26.

The thought that it might be worth trying to put together in a more systematic treatment the ideas that are in these earlier essays, suitably adjusted, emerged from my experience of trying to explain these ideas in a seminar sponsored by Philip Pettit at Princeton University's Center for Human Values. My recent thinking about these matters has benefited from discussions with very many people; but let me mention in particular discussions with Scott Shapiro when we both had the privilege of being Fellows at the Center for Advanced Study in the Behavioral Sciences in 2003-2004 while working on a joint project on shared agency, and discussions with Facundo Alonso and Margaret Gilbert. I am also much indebted to Alonso, and to Luca Ferrero and Abraham Roth, for detailed and insightful written comments on earlier drafts of this manuscript. And I have benefited greatly from a pair of fellowships at the Stanford Humanities Center and, in the early years of this work on shared agency, support from Stanford's Center for the Study of Language and Information and a fellowship at the Center for Advanced Study in the Behavioral Sciences. I am deeply grateful to these many people and institutions for their invaluable aid and support.

Table of Contents

7. The compressed basic thesis
8. Too demanding?

Chapter Five: Modest Sociality and Strategic Interaction

1. Modest sociality, intention and expectation
2. Modest sociality, intention and evaluation
3. Modest sociality and inter-connected planning agency
4. Taking stock

Chapter Six: Modest Sociality, Mutual Obligation

1. Social ties, explanation, and social rationality without mutual obligation
2. Shared intention, Moral obligation
3. Alonso on Shared Intention, Reliance, and Obligation
4. Gilbert on joint commitment
5. Normativity, sociality, and Ockham's Razor

Chapter Seven: Group Agents Without Group Subjects

1. Group agents and the basic thesis
2. Group subjects?

Chapter Eight: Shared Deliberation and Shared Policies of Acceptance

1. Two aspects of shared deliberation
2. Shared commitments to weights
3. Shared policies about weights
4. Interdependence in policies about weights
5. Partiality and depth of shared policies about weights
6. Shared policies of acceptance
7. Shared policies of social rationality

Conclusion: Interconnected Planning Agents

Chapter One: Sociality and Planning Agency

1. Modest sociality and the continuity thesis

Human beings act together in characteristic ways. Forms of shared activity matter to us a great deal, both intrinsically – think of friendship and love, singing duets, dancing together, and the joys of conversation -- and instrumentally – think of how we frequently manage to work together to achieve complex goals. Such forms of sociality are deeply involved in our lives. A human life that was not significantly embedded in such shared agency would probably be unrecognizable by us, and certainly be impoverished. The very phenomenon of language is tied in fundamental ways to shared activities of communication and the like. And, indeed, some have conjectured that our capacities for certain forms of shared activity set us apart as a species.[2]

In short, such shared activities are a fundamental element of a human life. For this reason such shared activities are – or, anyway, should be – a central concern of research in the social sciences, and an important target for philosophical and humanistic reflection. And my project in this book is to reflect on such basic forms of sociality: What concepts do we need to understand them adequately? In what do they consist? How are they related to relevant forms of individual agency? How does such sociality fit into the natural world? What norms are central to such sociality, and how are these norms related to norms that apply in the first instance to individual agency?

My pursuit in this book of these and related questions has a trio of inter-related aims, an underlying conjecture, and an important limitation.

The first aim is in the tradition of philosophy as conceptual articulation and innovation. The aim is to provide elements of a sufficiently clear and articulated framework of ideas to help support careful theorizing about these basic forms of sociality both in philosophy and in the wide range of other domains and disciplines within which these phenomena are of significance.

---

[2] See Michael Tomasello et. al., "Understanding and sharing intentions: The origins of cultural cognition," Behavioral and Brain Sciences 28 (2005): 675-735; Henrike Moll and Michael Tomasello, "Co-operation and Human Cognition: The Vygotskian Intelligence Hypothesis," (in press) Philosophical Transactions of The Royal Society B. Michael Tomasello and Malinda Carpenter, "Shared intentionality," Developmental Science 10:1 (2007), 121-125

The second aim is in the tradition of philosophical concerns with the metaphysics of human agency and its place in the natural world. Just as we can wonder about the place of individual agents in a natural causal order, we can ask about the nature of social agency and its place in that natural causal order. Just as we can ask how individual agents can be a source of their intentional activities in a natural causal order,[3] we can ask how we can be a source of our intentional activities in that causal order. And my aim is to approach these metaphysical questions about our sociality by exploring the continuities (if such there be) between the metaphysics of individual planning agency and the metaphysics of shared agency.

Finally, my third aim is in the tradition of normative philosophy. The aim is to understand the relation between central norms characteristic of shared agency and central norms that apply to individual agency. In particular, I want to know to what extent these social norms are grounded in norms of individual rationality.

In pursuit of all three aims a central idea is that a rich account of individual planning agency facilitates the step to basic forms of sociality. We begin with an underlying model of individual planning agency, one that I have called the planning theory.[4] And my guiding conjecture is that such individual planning agency brings with it sufficiently rich structures – conceptual, metaphysical, and normative – that the further step to basic forms of sociality, while deeply significant and demanding, need not involve fundamentally new elements. There is here a deep continuity between individual and social agency. This is an aspect of the fecundity of planning structures, the idea that planning structures in an important sense ground a wide range of fundamental practical capacities that are central to our human lives.

I do not claim that planning agency by itself ensures the capacity for the important forms of sociality that will be our focus. There will be theoretical room for planning agents who do not yet have the capacity for participation in the kinds of sociality highlighted here. The claim is only that once such individual planning agency is on board, what further is needed to make the step to such sociality is – in a sense to be explained -- conceptually, metaphysically, and normatively conservative. Nor, of

[3] This is a focus of my Structures of Agency

[4] See my Intention, Plans, and Practical Reason (Cambridge: Harvard University Press, 1987; re-issued by CSLI Press, 1999), and my Faces of Intention.

course, do I claim that appeal to planning structures captures all of the stunning practical complexity of human agents. We have complex and, at times, opaque emotional lives. We are prey to many different motivational pressures. Self-understanding is in many cases difficult. Wholeheartedness in thought and action can be elusive.[5] And much more could be said about our practical complexities and, as Jennifer Rosner has put it, our messiness.[6] My claim is only that planning structures are one salient and theoretically important aspect of the psychology that underlies our agency. In the case of individual agency such structures play a central role in important and characteristically human forms of cross-temporal organization. And my conjecture here is that versions of these structures are also an important part of basic forms of sociality.

This approach to our shared agency lies in the space between two important alternatives. The first is a commonsense version of the idea of Nash equilibrium in game theory. The idea is that in shared agency each is acting in pursuit of those things she wants or values in light of what she believes the other is doing, and where she knows the other's actions will depend in part on what the other thinks she will do. Each does best, by her lights, given what the other does. There is in that sense a strategic equilibrium. And all this is public, out in the open – where to be out in the open is, on one common interpretation, to be a matter of common knowledge.[7]

What does it mean to say that all this is common knowledge? There is a large and complex literature here, involving several different approaches.[8] One standard

[5] This is an important theme from Harry Frankfurt. See e.g., his "Identification and Wholeheartedness," as reprinted in his The Importance of What We Care About (Cambridge University Press, 1988).
[6] Jennifer Rosner, ed., The Messy Self (Boulder, Colorado: Paradigm Publishers, 2007).
[7] A recent treatment broadly in the spirit of this idea is Sara Rachel Chant and Zachary Ernst, "Group Intentions as Equilibria," Philosophical Studies 113 (2007): 95-109.
[8] Classic statements are in David Lewis, Convention: A Philosophical Study (Harvard University Press, 1969), 52-60; and Stephen Schiffer, Meaning (Oxford; Oxford University Press, 1972). See also Jon Barwise, "Three Views of Common Knowledge," in M. Vardi, ed., Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge (Los Altos, Ca.: Morgan Kaufman, 1988): 365-79; and Gilbert Harman, "Review of Linguistic Behavior by Jonathan Bennett," Language 53 (1977): 417-424, and "Self-reflexive Thoughts," Philosophical Issues 16 (2006): 334-345, esp. p. 342. For a useful overview see Peter Vandershraaf and Giacomo Sillari, "Common Knowledge," Stanford Encyclopedia of Philosophy. And see also Daniel Sperber and Diedre Wilson, Relevance (Harvard University Press, 1986), 38-45.

approach appeals to cognitive hierarchies.[9] One way to put this idea is to say that it is common knowledge among A and B that p just when (a) A knows that p, (b) B knows that p, (c) A knows that B knows that p, (d) B knows that A knows that p, (e) A is in an epistemic position to know that (d), (f) B is in an epistemic position to know that (c), and so on -- where once we get past (d) the stages of the hierarchy are a matter of what A and B are in an epistemic position to know, not of what they explicitly know.[10] And an underlying idea here is that common knowledge is a structure of inter-related cognitive aspects of the minds of relevant individuals.

Without trying to sort out the complex issues raised in the literature on common knowledge, I will mostly work with an intuitive notion of common knowledge. However, when it is useful to have a more specific model of common knowledge on hand, I will appeal to a hierarchical model along the lines just sketched. My hope, though, is that the main points I want to make are available to alternative treatments of common knowledge.

This gloss on the idea of common knowledge in hand, return to the cited model of shared activity as a strategic equilibrium within such common knowledge. What should we say about such a model? Well, I think there is a fundamental problem with this approach to our shared agency. The problem is that such mutual adjustment of each to another that is in a context of common knowledge and in strategic equilibrium – while an important phenomenon -- seems not by itself to ensure the kind of sociality we are after. When two strangers walk alongside each other down a crowded Fifth Avenue without bumping into each other, their patterns of walking near each other might be in strategic equilibrium in a context of common knowledge: each is acting in pursuit of what she wants in light of her beliefs about how the other is and will be acting, where what the other does depends on his beliefs about what she will do, and all this is out in the open. Each might in this sense be acting strategically in the light of what she values, the expected actions of the other, expectations about the corresponding expectations of the other, and so on. Yet they still might not be engaged together in a

[9] An alternative is a "fixed point" account sketched by Gilbert Harman and developed by Jon Barwise. (See references in previous note.)

[10] As Joshua Cohen once put it, such common knowledge is, in this respect, more like a competence than a performance.

shared activity of the sort we are trying to understand – they might not be, in the relevant sense, walking together.[11]

Agents who are walking together, or singing a duet together, or painting a house together, or having a conversation together, or dancing together -- where these are cases of the kind of sociality of central interest here – will, we may assume, be capable of employing such strategic reasoning and arriving at such an equilibrium. My claim is only that we cannot adequately model the kind of shared activity in which we are here primarily interested as simply a matter of the deployment of such strategic reasoning and a resulting equilibrium, given relevant common knowledge.

This does not mean that the very idea of an equilibrium within common knowledge is not an important idea for understanding sociality. Indeed, the theory I will be developing provides a model of a distinctive kind of shared practical settled-ness within common knowledge.[12] The point is only that what we learn from this example of, as I will say, walking alongside a stranger, is that a strategic equilibrium of the sort just described is by itself too weak to ensure the kind of sociality at issue.

Granted, there can be more complex strategic equilibria than this simple one of walking alongside a stranger. Perhaps a boy and a girl on Fifth Avenue, while strangers in the night, each walk down the avenue in a way that aims at ensuring that he or she achieves his or her personal goal of remaining close to the other.[13] Nevertheless, they still might not be engaged in what it is natural to classify as a shared intentional activity of walking together.

I will return to these matters below.[14] The point now is that reflection on such cases may suggest an alternative approach to articulating what is special about walking together as a shared activity, in contrast with a case of walking alongside a stranger.

---

[11] See Margaret Gilbert, "Walking Together: A Paradigmatic Social Phenomenon," Midwest Studies 15 (1990): 1-14.

[12] The relevant settled-ness will involve relevant intentions of each of the participants. Since Nash equilibrium is normally characterized in terms of the preferences of the relevant agents, rather than in terms of the agents' intentions, and since on my view intentions are to be distinguished from preferences, the kind of settled-ness within common knowledge that I will be focusing on need not be, strictly speaking, a kind of Nash equilibrium. Nevertheless, there is a broader way in which the model I will be proposing involves an equilibrium within common knowledge

[13] This example was suggested by comments by Paul Weirich at the 2008 University of Missouri workshop on shared agency.

[14] Chapter 5.

This alternative approach highlights the ways in which participants in such sociality are under obligations each to the other and are entitled to hold each other accountable. The idea is that the interconnections characteristic of such sociality essentially involve mutual entitlements to hold the other accountable for playing his role in the shared activity, and mutual obligations, each to the other, that correspond to these entitlements. And if we understand these entitlements and obligations as not essentially moral, we arrive at an idea that is central to Margaret Gilbert's understanding of shared agency.[15]

Now, I agree that mutual obligations and entitlements are extremely common in cases of human shared agency, though I think that these will be familiar kinds of moral obligations – in particular, moral obligations associated with promises, assurance, and reliance, as well as just plain civility. But, first, I am not convinced such obligations are always present. Second, I think that such obligations need not ensure the basic psychological elements of shared agency, since it remains possible for the parties to have no intention to conform to these obligations. And third, I think there are significant resources – conceptual, metaphysical and normative – in the territory between these two models, and that we do well to see if there is a viable theory in that territory.

I will return to and try to defend these points in the discussion to follow. Here I just want to note that the idea of the fecundity of planning structures aims to provide resources for a view in the space between these two other approaches. It seeks a view of our shared agency that is stronger than the idea of a strategic equilibrium within common knowledge, but still does not see inter-personal normative relations of mutual obligation and accountability as at the heart of the phenomenon. The idea is that the model of individual agency provided by the planning theory puts us in a position to understand what is missing from cases of merely strategic interaction that is needed for shared agency; and it puts us in a position to do this without a ground-level appeal to inter-personal relations of mutual obligation and mutual accountability.

Those, anyway, are my three inter-related aims, and my underlying conjecture. The limitation is that my focus will be primarily on the shared intentional activities of small, adult groups in the absence of asymmetric authority relations within those

[15] See e.g., Margaret Gilbert, "Shared Intention and Personal Intentions," Philosophical Studies 144 (2009): 167-87.

groups, and in which the individuals who are participants remain constant over time. Further, I will bracket complexities introduced by the inclusion of the group within a specific legal institution such as marriage, or incorporation. My interest will be primarily with duets and quartets rather than symphony orchestras with conductors, with small teams of builders rather than large and hierarchical construction companies, with small neighborhood groups rather than county governments, with small group discussion rather than deliberations in the US Senate, and with friendship rather than legally constituted marriage. And I will assume that these small groups have a stable membership.

I do not deny that there is an important sense in which there are larger institutional agents like corporations or governments, institutions with hierarchical authority relations, with potential flux in the list of their members, and, perhaps, with an embedded distinction between those participants who are officials of the institution and those who are not.[16] Rather, I hope to gain some insight by focusing initially on the kind of small-scaled shared agency to which I have pointed. Perhaps our theory of small-scale shared agency can, with due adjustment and further additions, be extended to such larger social organizations.[17] But first things first. I will be satisfied here if we can agree on a basic approach to the indicated kind of small-scale case of shared agency – as I will say, the case of modest sociality -- and leave to other occasions these potentially important extensions.

And my conjecture is that the conceptual, metaphysical, and normative structures central to such modest sociality are – in a sense I aim to explain -- continuous with structures of individual planning agency.[18] This is the continuity thesis. As we might try

[16] For this last distinction see H.L.A. Hart, The Concept of Law.

[17] For an effort to do this in understanding law see Scott Shapiro, Legality (Harvard University Press, 2011). For some caveats see Matthew Noah Smith, "The Law as a Social Practice," Legal Theory 12 (2006), 265-92. In her Liberal Loyalty: Freedom, Obligation, and the State (Princeton University Press, 2009) Chapter 7, Anna Stilz seeks an extension to the theory of democracy. For a systematic effort to extend an account of small-scale shared agency to larger institutions see Seamus Miller, Social Action: A Teleological Account (Cambridge: Cambridge University Press, 2001).

[18] Sophisticated and complex individual planning agency will, I take it, involve language; but perhaps simpler forms of planning agency need not essentially involve language, though language capacities, if present, would be supportive; and similarly concerning modest sociality. There are large questions and much debate in cognitive science about the relation between various forms of thinking and language. My hope is to develop an approach to shared agency that would be broadly available to a wide range of views about this matter. For a helpful overview and discussion of current work on the role of natural

saying: once God created individual planning agents and placed those agents in a world in which they have relevant knowledge of each other's minds, nothing fundamentally new -- conceptually, metaphysically, or normatively -- needs to be added for there to be modest sociality.[19] This is because the further steps from individual planning agents who know about each other's minds, to modest sociality, while substantive and demanding steps, are nevertheless primarily applications of the conceptual, metaphysical, and normative resources already available within our theory of individual planning agency. So, on this view, the problem of how our modest sociality is located in the natural world reduces to the problem of how our individual planning agency, in a context of knowledge of the minds of others, is located in the natural world.

In saying this I am assuming that the conceptual, metaphysical, and normative resources needed to model relevant forms of common knowledge are available within our model of individual planning agency. However, since, as noted, I do not try to defend a specific theory of the nature of common knowledge, this assumption is not fully defended here. If it were to turn out that this assumption is incorrect – that some further, fundamental resource is needed – then we would need to qualify the continuity thesis accordingly. But such a qualified continuity thesis would still maintain what is the central point here, namely that the theory of individual planning agency puts us in a position to provide a model of modest sociality without the introduction of fundamentally new practical elements that go beyond whatever cognitive structures are involved in common knowledge.

Let me try briefly to contrast this continuity thesis – whether qualified or not -- with approaches taken in work of John Searle and Margaret Gilbert, work I will be discussing in more detail below. The crucial point here is that both of these philosophers see the step from individual to shared agency as involving a new practical,

---

language in human thinking see Peter Caruthers, "Language in Cognition," in E. Margolis, R. Samuels, and S. Stich (eds.), The Oxford Handbook of Philosophy of Cognitive Science (Oxford University Press, 2011).

[19] Cp. Saul Kripke, Naming and Necessity at 153. While this thought experiment helps to point to the kind of continuities with which I am concerned, I make no claims about the actual etiology of our sociality. As noted, this conjecture leaves room for actual creatures who are planning agents with knowledge of the minds of others but who do not have the further capacities involved in modest sociality. The continuity thesis is not that no further capacities are needed for the step from planning agency to modest sociality, but rather that these further capacities need not involve fundamentally new elements.

metaphysical resource. In Searle's view, and as we will see in more detail below, what is needed is a new attitude of "we-intention".[20] In Gilbert's view, and as we will see in more detail below, what is needed is a new relation of "joint commitment" between the participants, a relation that necessitates distinctive mutual obligations.[21] In each case the cited new element is a practical element that is not just a matter of common knowledge. And both philosophers then try to understand larger institutions in ways that draw substantially on the new element that they cite as central to small-scale shared agency.[22] In contrast, my approach begins by distinguishing, in the individual case, between simple goal-directed agency and planning agency. Once individual planning agency is on board, the step to modest sociality need not involve a fundamental discontinuity[23] – which is not to say that all planning agents have the capacity for such sociality.[24] But this planning approach leaves it open how best to move from a theory of modest sociality to a theory of larger institutions.

2. Shared Intention

Suppose then that you and I are painting a house together. What makes this a shared intentional activity? We could imagine a contrast case in which we each intentionally go through the same motions as we do when we paint the house together, and yet there is no shared intentional activity. Perhaps we are each set only on our individual painting project and respond to each other only with an eye to avoiding collisions. While each of us acts, and acts intentionally, in a context that involves the other, ours is not a shared intentional activity: we are only painting alongside each other. Echoing Wittgenstein's question about the difference, in the individual case, between my arm's rising and my raising it,[25] we can ask: what is the difference between

[20] And Searle insists that "we-intentions are a primitive phenomenon". See John R. Searle, "Collective Intentions and Actions," in P. Cohen, J. Morgan, and M.E. Pollack, eds., Intentions in Communication (Cambridge: MIT Press, 1990): 401-415, at 406. Searle later expresses a version of this as the view that "the forms of collective intentionality cannot be eliminated or reduced to something else." John R. Searle, The Construction of Social Reality (The Free Press, 1995) at p. 37.
[21] Margaret Gilbert, "Shared Intention and Personal Intentions".
[22] John R. Searle, The Construction of Social Reality; Margaret Gilbert, A Theory of Political Obligation (Oxford University Press, 2006).
[23] Given my cited assumption about common knowledge.
[24] Indeed, a central conjecture of Michael Tomasello's is that the great apes are planning agents who know about the minds of other apes but do not have a capacity for shared intentional activity. See Michael Tomasello, Why We Cooperate (MIT Press, 2009).
[25] Philosophical Investigations par. 621; though Wittgenstein himself was wary of this question.

such a contrast case and corresponding shared intentional activity? In the case of individual intentional human action we can see the difference from a contrast case as involving an explanatory role of relevant intentions of the individual agent.[26] As a first approximation, I propose an analogous view of the shared case: the difference in the case of shared agency involves an appropriate explanatory role of relevant shared intentions. Our painting together is a shared intentional activity, roughly, when we paint together because we share an intention so to act.

Granted, there are also phenomena of group or collective agency, broadly construed, within which the distinctive social organization is grounded in causal mechanisms of a very different sort than that of shared intention. These will be forms of joint or collective agency in a weak sense, a sense does not involve shared intention. A swarm of bees, or a flock of birds, acts as a unit, yet ideas of intention – individual or shared – are unlikely to get at the relevant explanatory mechanisms.[27] And a panicky human crowd might cause the collapse of the bridge in the absence of any associated shared intention. I will return to such cases below. Here my conjecture is only that in central cases of human shared activity – cases naturally described as ones of shared intentional activity -- the concept of shared intention does point to important explanatory structures, and that it is these explanatory structures that are central to our answer to our social analogue of Wittgenstein's question.

What then is shared intention? And what is an appropriate explanatory relation? These questions will occupy me throughout this book, but let me begin here with some main ideas. In this first chapter I will focus primarily on the first question, what is shared intention? In chapter two I focus both on shared intention and on the cited explanatory relation (as I call it, the connection condition).

As I have indicated, my approach to shared intention is part of an effort to forge a path between two extremes: a model of strategic equilibrium within common knowledge, and a model of distinctive inter-personal obligations and entitlements to hold accountable. The middle path I will try to forge is an augmented individualism. It is an

[26] Though, as I have argued elsewhere, it is not true quite generally that when I A intentionally my action is explained by my intention specifically to A. See Intention, Plans, and Practical Reason (Cambridge: Harvard University Press, 1987; re-issued by CSLI Press, 1999), chap. 8.
[27] The example of a swarm of bees comes from Bjorn Petersson, "Collectivity and Circularity," Journal of Philosophy 104 (2007): 138-56. I will return to this essay below in chapters 5 and 7.

individualism that builds, in particular, on a rich story of our individual planning agency, one that goes beyond the desire-belief model in philosophy[28] and the associated utility-probability model in some areas of social science. And it is a story that augments the model of individual planning agency by highlighting special contents of and inter-relations between the plan-states of such individual agents.

Such an augmented individualism is not dismissive or debunking of phenomena of modest sociality. Far from it. The idea, rather, is to highlight the significance of these social phenomena and to provide a theoretically fecund structure of conceptual, metaphysical and normative resources in terms of which we can more deeply understand these phenomena.

This approach involves taking the intentions of individuals seriously as basic and distinctive elements of individual human agency, elements that go beyond the ordinary desires and beliefs[29] characteristic of simple purposive agency. Such intentions are embedded in coordinating plans that play basic roles in the temporally extended structures that are characteristic of human agency and that are in an important sense guided by associated norms of plan rationality. This is the planning theory of the intentions of individuals.

The central idea is that intentions are plan states. Such plan states are related to but different from ordinary beliefs and ordinary evaluations. First, intentions are not merely ordinary beliefs about one's own present or future conduct. After all, knowing myself as I do, I might now be confident that when I am this evening faced with the temptation of a second glass of wine at dinner I will give in to that temptation; and yet I might still not now intend to drink that second glass of wine this evening.[30] Again, I might believe that in intentionally X-ing I will produce a certain causal upshot, Y, without intending to produce Y. (A much discussed example: A bomber intends to destroy a

[28] A classic statement of this desire-belief model is in Donald Davidson, "Actions, Reasons, and Causes," in his Essays on Actions and Events (Oxford: Oxford University Press, 1980). A more recent defender is Michael Smith, The Moral Problem (Oxford: Blackwell, 1994), chap. 4. In a later paper, "Intending," (also in his Essays on Actions and Events) Davidson tried to move somewhat away from a simple desire-belief theory. I argue that his efforts did not fully come to terms with the role of intention in planning in my "Davidson's Theory of Intention," reprinted in my Faces of Intention.

[29] Or, in the case of fairly simple agents, conative and cognitive analogues of desire and belief.

[30] Anscombe gives a more extreme example: "a man could be as certain as possible that he will break down under torture, and yet determined not to break down." G.E.M. Anscombe, Intention second edition (Ithaca: Cornell University Press, 1963), 94.

weapons factory in order to promote the war effort, expects thereby to destroy a nearby school, but does not intend to destroy the school. After all, he would not go back and try again if somehow the school remained intact despite the destruction of the factory.)[31]

Turn now to the relation between intention and evaluation. Intentions will normally conform to the agent's judgments about what would be best, or the agent's rankings of options from best to worst. But intentions are not to be identified with such judgments or rankings. There are many cases of intending A while judging that an alternative B is as good, or ranking B as high in one's evaluative ranking, or just being unsure about which option is best. In the face of such under-determination by value judgment or value ranking one's intention is a form of selection of a specific option, a form of selection that allows us to avoid being human versions of Buridan's infamous ass. Further, there are also cases of judging B best and yet still being undecided. And there are cases of weak-willed intentions that are counter to what one judges best or ranks highest in one's evaluative ranking.

These reflections suggest that we best understand intention not as ordinary belief or evaluative judgment[32] or evaluative ranking, but rather in terms of characteristic roles – and associated norms -- in our planning agency.

I will say more about these distinctive, plan-related features of intention below. For now let me emphasize a general conjecture that is in the background. The conjecture is that a theoretically fruitful strategy in the philosophy of action is to try to understand important aspects of our agency by building on the planning theory. This is the idea, noted earlier, of the fecundity of planning structures, the idea that a sufficiently rich model of the planning structures that are common in individual human agency helps

[31] In claiming that intentions are not ordinary beliefs I put aside theories that maintain that intentions are a special, distinctive kind of belief. For example, in different ways both David Velleman and Kieran Setiya see an intention to A as a belief that one will A because one so believes, where this reflexive belief does indeed have this causal tendency. (However, they give different explanations of this causal tendency.) [J. David Velleman, … For clarification and adjustment see Velleman's "Introduction" to the 2007 re-issue of Practical Reflection at p. xix. , Kieran Setiya, .. .] I have discussed and criticized such theories elsewhere. [See, for example, my "Intention, Belief, Practical, Theoretical".] For now it suffices to make the simpler point about the contrast between intention and ordinary belief.

[32] In his "Intending" Donald Davidson tries to understanding intending as such an evaluation. I criticize this view in my "Davidson's Theory of Intention," as reprinted in my Faces of Intention.

illuminate other important aspects of human agency.[33] These further aspects of our agency include our capacity for complex, temporally extended activity, our capacity for self-governance, and our capacity for sociality. I have elsewhere explored this strategy as an approach to temporally extended agency and individual self-governance.[34] Here my primary target is modest sociality.[35] In the background is the idea that the step from desire-belief to planning structures, while of great significance, need not introduce any new, additional obstacles to locating our agency in the natural order. In this respect this step contrasts with appeals to forms of uncaused willing or of agent causation that are seen as outside the natural causal order and thereby threaten a broadly naturalistic conception of our agency.[36]

When, in light of this strategy, I turn in particular to shared intention, I try to see it as consisting in relevant plan-embedded intentions of each of the individual participants, in a suitable context and suitably inter-connected. As we might say: the shared-ness or joint-ness of shared intention consists in relevant contents of the plan states of each and relevant interconnections between the planning psychologies of each, all in relevant contexts. This augmented individualism depends then on a rich model of the individual agent as a planning agent whose agency is temporally extended.

Granted, much work in the philosophy of mind has argued that our ordinary ways of specifying the contents of the attitudes sometimes draw on features outside of the individual whose attitudes are at issue. These external features may include the causal context of the use of natural kind terms or of names[37] and/or relevant linguistic practices.[38] However, the kind of augmented individualism I seek is officially neutral about these debates about the nature of content. What is crucial to this augmented individualism is not whether the contents of the attitudes of the individuals involve appeal to elements outside those individuals. What is crucial, rather, is that shared

[33] I also discuss this idea in my "Agency, Time, and Sociality," Proceedings of the American Philosophical Association.
[34] The main target of Intention, Plans, and Practical Reason, was our temporally extended agency. The main target of my Structures of Agency was our individual self-governance.
[35] Shapiro's approach to the law is in this spirit. As he says "we are able to create law because we are able to create and share plans." (Legality,181).
[36] So the planning theory can be seen as a de-mystifying theory of the will.
[37] Hilary Putnam, "The Meaning of 'Meaning'," in his Mind, Language and Reality (Cambridge: Cambridge University Press, 1975), vol. 2.
[38] Tyler Burge, "Individualism and the Mental," Midwest Studies in Philosophy 4 (1979): 73-121.

intention consists primarily of relevant inter-related attitudes of these individuals (especially, intentions), and that the contents of the attitudes that are constitutive of basic cases of shared intention need not (though on occasion they may) essentially involve the very idea of shared intention.

In developing this idea, what we will see is that though the contents of the intentions of each that are central to shared intention need not involve the very idea of shared intention, they will nevertheless have a distinctive character. In particular, in contrast with ordinary cases of intending to do something oneself, shared intention involves intentions of the individuals whose contents appeal to the group activity. Our shared intention to paint together involves your intention that we paint and my intention that we paint. Or anyway, that will be a central feature of the theory I will be developing.

This violates the own-action condition on the content of intention. According to this own-action condition it is always true that the subject of an intention is the intended agent of the intended activity. And it does seem initially plausible that intentions should respect some such constraint. Just what lies behind this initial intuition -- and what its force is, on reflection -- is a matter to which I will return in chapter 3. Here I just want to acknowledge that the view I will be developing does involve rejecting this own-action condition, since it appeals to intentions of each individual participant that they (the group) act.

The apparent problem here is not initially a problem for talk of our intention to do something together. After all, when I say that we intend to paint together the intention I report is our shared intention in favor of our shared action.[39] But my proposal is to understand shared intention by appeal, inter alia, to my intention that we paint. Since that violates the own action condition something needs to be said.[40]

One reaction to this is in the spirit of work of John Searle.[41] As noted, Searle focuses on what he calls "we-intention". What he means by this is not what I mean in talking about our intention. As I see it, the fact that we intend to paint involves both me,

[39] But these matters get more complex upon further analysis. See chapter 7.
[40] Versions of this worry, addressed to my initial essays on this subject, were voiced by Annette Baier, Frederick Stoutland, J. David Velleman. For references, and my earlier response to these challenges, see my "I Intend that We J," as reprinted in my Faces of Intention. And see also Neil Roughley's review of my Faces of Intention in International Journal of Philosophical Studies 9 (2001), 265-270, at 268-9.
[41] "Collective Intentions and Actions," in P. Cohen, J. Morgan, and M.E. Pollack, eds., Intentions in Communication (Cambridge: MIT Press, 1990): 401-415.

you, and relations between us. In contrast, Searle's we-intentions are attitudes in the head of an individual, though attitudes that concern the activity of a "we". You could have a we-intention, in Searle's sense, if you were in fact the only person in the world, but thought there were others with whom you might, say, dance. A Searlean we-intention is, then, a candidate for the intentions of individual participants that together help constitute a shared intention, though Searle himself does not, to my knowledge, say how precisely the we-intentions of different participants need to be inter-related for there to be (in my sense) shared intention.

Searle's we-intentions violate the own-action condition. This may be at least one reason why he claims that we-intentions are not just ordinary intentions with a special content, a content that involves the activity of a "we". We-intentions are, rather, a special intending attitude, to be distinguished from the ordinary attitude of intending involved in individual agency.[42] If we suppose that the ordinary attitude of intending is subject to the own-action condition, and if we countenance we-intentions, then it will be natural to see we-intentions as distinctive attitudes rather than as ordinary intentions with a special content. In this way we can be led from reflection on the conceptual resources at work in the content of the attitude to a view about the metaphysics of mind, a view that posits a distinctive attitude of we-intention and a corresponding breakdown in the continuity thesis.

An alternative approach is to emphasize the commonalities in the attitudes involved in intending to act and intending that we act, and to see the differences as deriving from differences in content and distinctive features of the first-person singular as it appears in certain of these contents. Just as I can believe that I will do something, I can also believe that we will do something: in both cases what I have is an ordinary belief. On this alternative approach the situation is similar in the case of intention. This

[42] Or anyway, that seems to me the best reading of Searle's essay. As noted earlier, he explicitly asserts that "we-intentions are a primitive phenomenon".(406) And he says that "we-intentions are a primitive form of intentionality".(407) He then goes on to say that we-intentions are "not reducible to I-intentions plus mutual beliefs." (p. 407) This might suggest that he is only rejecting a specific reduction of "we-intentions", one he attributes in this essay to an early essay of Tuomela and Miller. But when Searle sketches his own positive theory his formal apparatus explicitly distinguishes, among intentions in action, those involved in individual intentional activity, and "collective" intentions in action; and this seems to embed a flat-out claim of non-reducibility. (see p. 412) For a study of this aspect of Searle's views see Kirk Ludwig, "Foundations of Social Reality in Collective Intentional Behavior," in S.L. Tsohatzidis, ed., Intentional Acts and Institutional Facts (Springer, 2007): 49-71.

approach allows us to draw directly on what the planning theory tells us about the nature of ordinary intention. In contrast, this is apparently blocked by Searle' strategy, since Searle's we-intentions are not themselves ordinary intentions.

In response, might Searle insist that, though "we-intentions are a primitive phenomenon," there is a kind of isomorphism between we-intentions and the intentions described by the planning theory, an isomorphism that allows us to understand we-intentions using resources from the planning theory?[43] Could Searle say that this is why we-intentions function in ways analogous to those specified by the planning theory? Well, on the one hand, insofar as we hold onto the view that we-intentions are fundamentally different from the attitudes described by the planning theory, it is not clear what gives us the right to assume that there is an isomorphism that can play this strong theoretical role. On the other hand, suppose that we can support the idea that we-intentions function in ways that are isomorphic to the ways in which intentions function within the planning theory. But then why not say – as I am saying -- that we-intentions are indeed intentions of the sort described by the planning theory, though intentions whose contents differ from ordinary intentions to perform one's own action? The distinction is not between two fundamentally different attitudes, but between two different kinds of contents of the attitude of intending, an attitude described by the planning theory.[44]

In any case, this latter conservative resistance against positing fundamentally new and primitive attitudes will be characteristic of the approach I will be taking. My proposal will be that we locate both intending to act and intending that we act in largely similar ways within the nexus of roles and norms highlighted by the planning theory of intention (roles and norms to be discussed below). Both intending to act and intending that we act are to be distinguished in this way from mere desires, or beliefs, or hopes, or judgments about what is best. In short, both intending to act and intending that we act are forms of intending, as that is characterized by the planning theory – though of

[43] This is my reconstruction of Searle's appeal to isomorphism in his informal response to my presentation at the Konstanz Conference, June 2011.

[44] We can also ask how Searle's approach avoids a proliferation of different attitudes. A single person will join with many different groups to form many different "we's". A might sing with B, bake with C, walk with D, and so on. Presumably, A does not need an AB-intention in favor of singing, an AC-intention in favor of baking, and an AD-intention in favor of walking.

course they also have different contents and associated differences in role. Both intending to act and intending that we act play plan-theoretic roles and are subject to associated norms of plan rationality, to be discussed below. And, in the end, we should judge this more conservative strategy – in contrast with Searle's appeal to a primitive phenomenon of we-intention -- by seeing how it contributes to the explanation and understanding of basic forms of sociality: the proof of the pudding is in the eating.

That said, there may remain a lingering sense that there is something odd in this talk of intending that we J. I will try to address these lingering worries below, in Chapter 3. But first I need to say more about my overall model of shared intention; and to do that I need to begin with the intentions of individuals.

3. Individual planning agency, and two faces of intention rationality

Given the importance of the idea of the fecundity of planning structures to my approach to shared agency, I need to explain in more detail how I will be understanding the intentions and plans of individual agents. In intending X or planning X the agent is in a distinctive way settled on X. As I have indicated, I do not think intending X is simply believing X or judging X best. But how then should we understand what is involved in being settled in this characteristic way?

Well, a basic idea, one that applies broadly in the philosophy of mind and action, is that we make progress in understanding aspects of mind by articulating characteristic functions or roles together with associated norms. This idea lies behind the planning theory of the intentions of individuals.

The planning theory sees intentions as plan states: they are embedded in forms of planning central to our temporally extended agency and to our associated abilities to achieve complex goals across time and inter-personally, even given our cognitive limitations. The theory appeals to the guiding, coordinating, organizing, and selecting roles of intentions as elements of larger – and, typically, partial, hierarchical, and future-directed – plans. And the theory sees these roles as supporting a basic distinction between intending and planning on the one hand, and, on the other hand, ordinary belief or evaluative judgment or desire or preference

Associated with this web of plan-like roles are characteristic norms, ones plausibly called norms of intention rationality. Primary among these norms are norms of

consistency, agglomeration, means-end coherence and stability: Intentions are to be internally consistent, and consistent with one's beliefs; and it should be possible to agglomerate one's various intentions into a larger intention that is consistent in these ways. Intentions in favor of ends engage demands, roughly, to settle, as needed and in a timely way, on means and the like – demands of means-end coherence. And while intentions are subject to reconsideration and revision, they are also subject to pressures in favor of stability over time.[45]

These norms of intention rationality are associated with the characteristic roles of intention in two inter-related ways. The first is explanatory: these norms will enter indirectly into a standard explanation of the normal ways in which these plan states play these roles in planning agency. In particular, we can suppose that in the case of individual planning agency the standard way in which these norms enter into such explanations is that their at-least-implicit acceptance by those planning agents helps explain how these plan states play these roles.[46] For example, intentions in favor of ends tend to issue in reasoning that aims at settling on means, and this is in part because the agent's associated thinking is guided by an accepted norm of means-end coherence. And intentions concerning the future tend stably to structure thought and action over time, and this is in part because of an accepted norm in favor of intention stability.

This idea of an accepted norm is tied to the idea of a disposition to see divergence from the norm as a mistake, a breakdown. If, for example, a planning agent realizes that her intention to A and her intention to B are not co-realizable, she will

[45] See my Intention, Plans, and Practical Reason. My formulation of the agglomeration principle here has benefited from the discussion in Gideon Yaffe, "Trying, Intending and Attempted Crimes," Philosophical Topics 32 (2004): 505-532, at 510-522. The temporal qualification in the norm of means-end coherence is important, since it is normally rationally permissible to leave means-end gaps in one's future-directed plans if one supposes that there will be time later to fill in those plans. This contrasts with the consistency norm: a present inconsistency in plans for the future will violate that norm even if there remains time to sort this out before the need to act. I discuss the nature and ground of these norms on intention further in "Intention, Belief, Practical, Theoretical," in Jens Timmerman, John Skorupski, and Simon Robertson, eds., Spheres of Reason (2009), "Intention, Belief, and Instrumental Rationality," in David Sobel and Steven Wall, eds., Reasons for Action (Cambridge University Press, 2009), and in "Intention, Practical Rationality, and Self-Governance," Ethics (2009). In these essays I argue, in particular, that these norms are best seen as distinctively practical norms, rather than as, at bottom, theoretical norms on the beliefs that are purportedly involved in intending.

[46] In talking about the individual agent's acceptance of these norms I lean on Allan Gibbard's extensive and philosophically powerful use of this idea in his Wise Choices, Apt Feelings. See also Peter Railton, "Normative Guidance," in Russ Shafer-Landau, ed., Oxford Studies in Meta-Ethics 1 (2006): 3-33.

normally think she has made a mistake. A common manifestation of this will be a kind of "darn it!" reaction. And this reaction will tend to lead to efforts to revise so as newly to come into relevant conformity with the norm.

A second way in which these norms of intention rationality are associated with the cited roles of intention involves the thought that these norms really do have normative force or significance.[47] When we reflect on these norms we continue to accept them in part because we can see that they have some such normative force or significance; or so it seems. And this thought is part of an explanation of the stability under reflection of the planning system. The planning theory seeks to explain what, if anything, gives these characteristic norms their distinctive force or significance. In this way the planning theory draws on substantive normative reflection and argument.

As I see it, a central part of this normative story will articulate the relation between general conformity to such norms and reliably realizing the characteristic coordinating, organizing, and selecting roles of planning; and it will explain what is good about reliably realizing these roles. The idea will be that being guided by one's at least implicit acceptance of these norms is an important element in how these roles are normally realized, and that it is important to us that these roles indeed be realized.

This is to explain the normative force of these norms in part by appeal to the importance to us of the general functioning their acceptance supports. We can also ask further whether there is something of non-instrumental significance in the satisfaction of these planning norms in each particular case to which they apply. And I have elsewhere defended a version of an affirmative answer to this question.[48] There are large and important issues here, but for present purposes we can rest content with a trio of ideas: First, the planning theory involves both a descriptive account of the underlying, accepted norms, and an account of the normative force or significance of those norms. Second, we can try to understand this normative force both by appeal to the importance of the general forms of functioning the acceptance of these norms supports, and in terms of the non-instrumental significance of the satisfaction of these

[47] This distinction between explanatory role and normative force is in the spirit of Timothy Schroeder's distinction between a "categorization scheme" of a norm and a "force-maker" of that norm. See Timothy Schroeder, "Donald Davidson's Theory of Mind is Non-Normative," Philosophers' Imprint 3 (2003): 1-14.
[48] See my "Intention, Practical Rationality, and Self-Governance," Ethics, and my "Why Be Means-End Coherent?" unpublished MS.

norms in the particular case. Third, given the way in which the acceptance of these norms is partly constitutive of planning agency, and the way in which their acceptance is central to the basic coordinating, organizing, and selecting roles of planning, we can see these norms as norms of plan rationality.

Might the planning theory instead retreat to a "positivistic" theory that simply describes how the planning economy works – and what norms are accepted by creatures who are characterized by such a planning economy – without a concern with whether those norms do indeed have normative force or significance? Well, since we who are theorizing about such planning agency are also ourselves planning agents, there is an instability for us in such a purely positivistic account. We ourselves, as planning agents, treat these underlying norms as having some sort of normative significance; indeed, we treat these norms in this way in the very planning that is involved in our efforts to construct such theories of planning agency. (Perhaps you intend to apply for an NSF grant as a means to your intended end of studying planning agency.) Since our theory of planning agency is a theory of our planning agency, there will be pressure on our theory to ask whether that treatment of these norms makes sense, and if so why. Further, given an account of the normative significance of these norms we will have reason to expect that these planning structures will in general be stable under reflection by planning agents like us.[49] And, as noted, such a view about the reflective stability of these planning structures will itself be an element in our overall theory of planning agency.

The planning theory, then, will appeal to characteristic roles of plan states, to an explanatory role of norms in explaining those roles, and, in particular, to forms of norm acceptance by individual planning agents that help explain those roles. And it will try to explain the normative force or significance of those characteristic norms. So, on the planning theory, the norms characteristic of planning agency have two inter-related faces: an explanatory face, and a normative face. In the case of individual planning agency the explanatory face of these norms consists in the ways in which their at-least-implicit acceptance by individual planning agents helps support characteristic planning roles. An account of the normative face of these norms -- an account of why it matters

[49] Which is not to say that we have it in our power simply to abandon our planning agency.

whether one conforms to these norms – will appeal in part to these explanatory roles of these norms and to the importance of the functioning that they thereby help support. And such an account of the normative force of these norms promises to help explain the reflective stability of these planning structures.

4. Individual planning agency: roles, norms, and Lockean ties

Let me now try to spell out in more detail some of the roles and norms that are, on the theory, characteristic of individual planning agency.

The central idea is that intentions are plan states, states that enter at least potentially into plans whose central role is coordination over time and, as we will be seeing, socially. Such plan states are not to be identified with predictions about how one will act. Nor are they to be identified with judgments of the best. In particular, one's intentions and plans many times involve a choice or selection of one of a number of alternatives each of which is seen by the agent as adequately supported by relevant considerations. This can happen is Buridan's ass cases, in which one forms an intention in the face of what one sees as equally desirable options. And it can happen in cases in which one sees several options as in important ways incomparable. Sartre's case of the boy who must decide between a life helping his mother and a life with the Free French is a classic example.[50]

Intentions, once they are formed or otherwise take shape in the mind of the agent, help constitute coordinating plans that are normally partial. Such partiality is central for agents who, like us, have significant limits of mental resources and significant limits in our abilities to predict the future. Given this characteristic partiality, such plans will need to be filled in as time goes by. In particular, as elements of such partial, coordinating plans, intentions pose problems of means and preliminary steps, problems that need to be solved in a timely way if one is to avoid means-end incoherence. And prior intentions constrain the formation of other intentions by filtering out of deliberation options whose performance would be inconsistent with one's prior intentions, given one's beliefs. Such prior intentions help provide continuity and organization over time, and, if all goes well, eventually control relevant conduct.

[50] Jean-Paul Sartre, "Existentialism Is a Humanism," in W. Kaufmann, ed., Existentialism from Dostoevsky to Sartre rev. and expanded (New York: Meridian/Penguin, 1975): 345-69, at 354-56.

Sometimes these plan states will have a certain generality: one can have an intention to buckle up one's seat-belts whenever one drives, or to have only one beer at dinners. These general intentions are policies. Such polices, while general in their content, will have implicit unless-clauses: my policy will probably not enjoin buckling up if I am in a certain kind of emergency situation.[51]

One can also have policies about what to treat as having weight in relevant deliberation.[52] The boy in Sartre's case might arrive at a policy of giving weight in his relevant deliberations to the political cause of the Free French. Such policies about weights will frequently be associated with judgments of value, but they need not correspond to a prior judgment of the best and may settle matters in response to under-determination by prior value judgment. This is what may happen in Sartre's case, or in less dramatic but more common career decisions.

As noted, in understanding these various kinds of plan states the planning theory appeals to characteristic norms of plan rationality, norms of consistency, agglomeration, means-end coherence and stability. These norms are normally internal in the sense that their – perhaps only implicit -- acceptance is normally operative in a planning agent's psychic economy and part of the explanation of the normal conformity to these norms.

Let me pause to reflect briefly on stability. The norm of stability in part concerns the reconsideration of intentions already formed. We normally retain our prior intentions unless we reconsider them. Reconsideration, however, takes time and uses other mental resources; and reconsideration may require, in the pursuit of coordination, re-thinking various other, related courses of action on which one had earlier settled. So there is frequently reason not to reconsider, both because of the direct costs of reconsideration and because of risks of undermining coordination previously forged. Further, so as not to use deliberative resources inefficiently we frequently depend on general, non-deliberative habits and strategies about when to reconsider. And habits and strategies that to some extent favor non-reconsideration will be likely, in the long run, to be conducive to the overall effectiveness of our temporally extended agency.

[51] See my "Intention and Personal Policies".
[52] I discuss the central role in self-governance of such policies about weights – as I call them, self-governing policies – in my "Three Theories of Self-Governance" as reprinted in Structures of Agency.

It is also true that once one has embarked on an intended course of action there will frequently be a snowball effect: things will frequently change in ways that support further new reasons to continue with what one intends.[53] One will, say, be closer to completing what one intended than one was before one began.

Once one does reconsider a prior intention, does the fact that up to now one has been so intending have its own normative significance? Well, a general habit or strategy of giving one's prior intentions a kind of default status in one's practical thinking seems likely to be broadly supportive of the temporal organization of our agency. I have also argued elsewhere – though this is controversial -- that it is also true that in the particular case the default stability of such a prior intention is normally an element in a kind of self-governance over time that we value.[54]

Returning to the broad nexus of roles and norms, a general idea, already noted, is that these roles and norms help distinguish intentions from the ordinary desires and non-evaluative beliefs characteristic of simple purposive agency.[55] There is much to be said here, but let me just briefly note some important differences. First, ordinary desires are not subject to the same norms of consistency and means-end coherence. It is, after all, part of the human condition to have desires for different things that, one knows, are not co-possible. And simply desiring something does not yet put me under rational pressure to ascertain or settle on means to it. Second, ordinary non-evaluative belief does not pose problems of means: simply believing that, given my social awkwardness, I will offend someone at the party does not pose a problem of settling on means to do that.[56] Indeed, even if I am taking steps aimed at preventing this upshot, I can still be resigned in my belief that I will nevertheless offend someone; whereas taking such preventative steps would not normally be compatible with intending to offend. Third, intentions play a distinctive filtering role. If I believe p but consider an option X that would prevent p, a decision to X would not thereby be blocked – I would just need to be prepared to update my belief about p were I to intend to X. But if I intend to bring about

[53] I owe the term "snowball effect" to John Etchemendy. See Intention, Plans, and Practical Reason at 82.
[54] See my "Agency, Time, and Sociality" and my "Self-Governance, Time, and Rationality" unpublished MS.
[55] And, as also noted, intentions also contrast with evaluative judgments.
[56] Granted, beliefs pose problems of coherence within one's beliefs. But I take it that means-end coherence is not reducible to belief coherence. See my "Intention, Belief, Practical, Theoretical".

p and then note an option Y that I know would prevent p, my prior intention would normally filter Y from consideration in my deliberation, and thereby block a decision to Y. And this filtering of options would normally be the expression of an underlying norm of intention consistency.[57]

We also want to distinguish intentions from the more general phenomenon of a goal. We can have, and act on, goals that we do not see as subject to a demand for agglomeration. Perhaps my goal in filling out a certain admissions form is to gain admission to Stanford Law School. When I turn to the form from UCLA my goal is gaining admission to UCLA Law School. But I know, let us suppose, that these law schools coordinate admissions and so it is not possible to gain admission to both law schools, though it is possible to gain admission to each. So I, quite sensibly, do not have – and indeed reject -- the goal of gaining admission to both. In this way goals differ from intentions, since intentions are, we have said, subject to a demand of being capable of being agglomerated without violating the consistency demand. Intentions are a special kind of goal state – namely, a plan state – and plan states are subject, in particular, to the cited demand for agglomeration and consistency.[58]

If all goes well, planning structures induce cross-temporal referential connections that are both forward and backward looking. My present plan to go to Boston next week at least implicitly refers to my later, then-present-directed intention to go by getting on the airplane; and my later intention at least implicitly refers back to my earlier intention. Further, the normal stability of such intentions over time helps support a coordinated flow of activity over time. These cross-temporal constancies and referential connections help support a temporally extended structure of partial plans that can provide a background framework for further deliberation aimed at filling in these plans as need be and as time goes by. And this further deliberation is shaped in part by rational pressures in the direction of means-end coherence, intention-belief consistency, and agglomeration. In these ways, a planning agent's purposive activity over time is

[57] These last two points are involved in the distinction, noted earlier, between intending A and believing one will A as a result of something else one intends. This distinction lies behind the moral principle of double effect. I discuss this distinction in Intention, Plans, and Practical Reason, chap. 10, and in "Intention, Belief, Practical, Theoretical".

[58] The idea that goals need not be subject to the agglomerativity constraints characteristic of intentions is a lesson of the video-games case I discuss in "Two Faces of Intention," and in Intention, Plans, and Practical Reason chap. 8.

typically embedded within interwoven structures of partial, referentially interlocking, hierarchical, and somewhat stable plan states, and in modes of further deliberation and planning that are motivated and framed by these plan states.

This idea of cross-temporally stable and referentially interlocking attitudes is familiar from the Lockean tradition of reflection on personal identity over time.[59] A central idea of that tradition has been that such identity over time essentially involves overlapping strands of continuities of attitude and broadly referential connections across attitudes.[60] And what we have seen is that the standard functioning in planning agency of attitudes of intending involves such Lockean cross-temporal ties.

The claim is not that the intention-like roles I have been highlighting are realized in all forms of agency, or that the associated norms on intention apply to all agents. After all, not all agents are planning agents. There can be purposive agents – dogs and cats, perhaps – who do not have the organizational resources of planning agency. But it seems plausible that we – adult humans in a broadly modern world -- are planning agents, and that this is central to characteristic forms of cross-temporal organization in our lives. The planning theory is a theory about the nature of intentions understood as central elements in this fundamental form of human, temporally extended agency. Such intentions bring with them a complex nexus of roles and norms that is characteristic of planning agency. And these structures go well beyond simple desire-belief purposive agency. So it seems reasonable to see intentions, so understood, as distinctive elements of the psychic economy of planning agency. This is the distinctiveness of intention.[61]

This emphasis on planning structures may seem to point to a distorted caricature of human agents as constantly planning, eschewing spontaneity, and rigidly following through with prior plans. And I agree there is a danger here of arriving at a one-sided

[59] John Locke, An Essay Concening Human Understanding Bk. 2, chap. 27. And see also H.P. Grice, Anthony Quinton, John Perry, and Derek Parfit, Reasons and Persons (Oxford: Oxford University Press, 1984).

[60] I discuss the relation between these comments and Derek Parfit's Lockean theory of personal identity in my "Reflection, Planning, and Temporally Extended Agency," as reprinted in my Structures of Agency, 21-46, at 28-30.

[61] This is a main theme in Intention, Plans, and Practical Reason. And this is why I reject the "strategy of extension" that I attribute to Donald Davidson. See my "Davidson's Theory of Intention" as reprinted in my Faces of Intention esp. 222-224.

picture of human agency. It is a remarkable fact about human agents that they have capacities that help to support and to constitute deep forms of cross-temporal and social coordination and organization. And it seems that planning capacities are central here. But of course these planning capacities are embedded in a complex psychic economy that also involves abilities to characterize one's plans in schematic and open ways,[62] and to be spontaneous and flexible as time goes by. A basic challenge for a theory of human agency will be to do justice both to the centrality of planning in the constitution and support of fundamental forms of organization and to our important capacities for conceptual openness, spontaneity, and flexibility. And here it will be natural to think about our agency in broadly virtue-theoretic ways, and appeal to relevant practical virtues that are involved in well-functioning planning agency.

The planning roles I am highlighting are primarily roles of intention in temporally downstream psychic functioning, including further reasoning and action. They are downstream roles in organizing, stabilizing, coordinating, and making effective our temporally extended activity. They are roles that involve characteristic forms of tracking and filtering; and in many of these temporally downstream roles prior intentions shape further, later deliberation – as when a prior intention in favor of E structures further reasoning about means to E. So a full story of these downstream roles will include a story of the roles of intentions in shaping such further deliberation.

Of course, intentions do not only shape further deliberation; they are also typically themselves an issue of prior deliberation. So a full theory of intention will also need to be in part a theory of the nature of the deliberation from which intentions are sometimes an issue. And indeed, I have argued elsewhere that our precise understanding of the way in which intention is an output of deliberation is central to our understanding of the important distinction between intending A and believing one will A as a result of something one intends.[63] Nevertheless, in its effort to say what intentions are, and how they are distinctive, the planning theory highlights in particular the

[62] This remark about conceptual openness owes to conversation with Keith Lehrer.

[63] In Intention, Plans, and Practical Reason chap. 10, I explore some of the complexities of putting together a story of the deliberation that leads to intention with my story of the downstream roles of intention. As I see it, this is a central issue that is raised by the question of what the difference is between intending A and believing one will A as a result of something one intends. And I argue that reflection on this issue leads us to distinguish intention from the holistic choice in which deliberation normally issues.

temporally downstream roles of intentions as elements in partial, coordinating plans – plans that serve, inter alia, as inputs to later practical reasoning. The step from simple purposive agency to planning agency is in large part a step to the capacity for attitudes that play these inter-related temporally downstream roles in organizing our temporally extended thought and action. And it is this capacity that will prove to be central to my understanding of shared agency.

5. Creature construction

We can see the step just described – the step from simple purposive agency to planning agency -- as a step in what Paul Grice calls "creature construction".[64] The aim of creature construction is to understand more complex forms of agency by building step-wise from simpler forms of agency. We build more complex structures upon a foundation of simpler structures in ways that respond to identifiable problems and issues that arise in the context of those simpler structures. And my proposal is, in particular, that we build structures of planning agency on top of structures of purposive agency in response to problems of coordination and organization both over time and, as we will see, socially. We do this in a way that is compatible with our cognitive limitations – where this includes limits on the time we have for reflection given the pressure for action, limits on the complexity of the contents of our thinking, and limits of knowledge about the future.[65] We do this in a way that is responsive to our need many times to chose or select among conflicting options in the face of under-determination of our choice by considerations about what it would be best to do. And we do this in a way that is responsive to our needs for self-control and self-management in the pursuit of organization and coordination and in the face of conflicting sources of motivation.[66]

---

[64] See Paul Grice, "Method in Philosophical Psychology (From the Banal to the Bizarre)," Proceedings and Addresses of the American Philosophical Association (1974-75): 23-53.

[65] See my "Valuing and the Will" as reprinted in Structures of Agency. Herbert Simon's work has been enormously influential in focusing our attention on the fundamental significance of these limits. See for example, his Reason in Human Affairs (Stanford: Stanford University Press, 1983). And for an application of some of these ideas to issues in artificial intelligence, see M. E. Bratman, D. J. Israel, and M E. Pollack. "Plans and resource-bounded practical reasoning," Computational Intelligence (1988): 349-355.

[66] Concerning the roles of intention-like attitudes in response to temptation see my "Toxin, Temptation, and the Stability of Intention," in Faces of Intention and my "Temptation Re-visited" in Structures of Agency. For the general theme of self-management see "Autonomy and Hierarchy," in Structures of Agency.

The idea is not that this is how our planning agency actually emerged within an evolutionary, historical process.[67] The idea is only that such a hypothetical series of constructed creatures can help us understand complex elements of our actual planning agency, elements that are compatible with our cognitive limitations and build on but go beyond less complex elements in ways that respond to basic concerns with cross-temporal and social coordination.

When we see planning agency as such a step in creature construction, it will be natural then to see the step from individual planning agency to shared agency as yet a further step in creature construction. And that is what I will do. Further, I will argue that this step to shared agency can be conceptually, metaphysically, and normatively more conservative than the step from individual purposive agency to individual planning agency.

But why build shared agency on top of, in particular, individual planning agency? Could there not be forms of joint or collective agency whose participants were simple purposive agents and not themselves planning agents? In alluding to swarms and flocks and crowds I have already granted that the answer to the second question is a qualified "yes". Nevertheless, my answer to the apparent challenge this poses is two-fold. First, we are interested in our shared agency, and this is shared agency whose participants are, in fact, planning agents. And second, the ability of the theory to refer to and exploit these planning structures allows it to provide a rich model of robust forms of shared agency without introducing fundamentally new and discontinuous elements. This is an aspect of the fecundity of planning structures, and supports the thesis of continuity. Or so I will argue.

6. Social Functioning and Social Rationality

We have available, then, a trio of guiding ideas. First, there is the general idea that we try to understand aspects of mind in terms of characteristic roles and associated norms. Second, we have available such an understanding of, in particular, the intentions of individuals as elements in partial, coordinating plans. This is the planning theory of individual agency. And third, we have available Grice's methodology of

[67] Indeed, as Mark Turner noted in conversation, the evolutionary story of our individual psychic economies may well need to appeal to ways in which they support sociality.

creature construction. Given this trio of ideas, how should we think about shared intention?

We can begin with our first, very general idea and apply it directly to shared intention. We ask: why do we bother with shared intentions? what fundamental roles do they play in our lives, and what norms are associated with those roles? Let's focus first on roles. And here I think we should be struck by the analogues, in the shared case, of the fundamental coordinating, structuring, and guiding roles of intention in the individual case. In particular, it seems plausible to suppose that the characteristic roles of a shared intention to J include the inter-personal coordination of action and planning in pursuit of J, and the structuring of related bargaining and shared deliberation concerning how to J.[68] In playing these roles shared intentions help to constitute and to support basic forms of social organization, forms that we value both intrinsically and instrumentally.

Granted, and as noted earlier, human shared agency many times also brings with it not only coordination of thought and action but also associated practices of holding accountable. As Margaret Gilbert has emphasized, one participant may well demand that the other do her share, and hold her accountable if she does not do this. But I think it is natural to see this not as a defining role of shared intention – as what shared intention is for – but rather as a frequent, supporting concomitant of human shared agency. The basic answer to why we bother with shared intentions and shared agency is not: so as to hold each other accountable. The more plausible answer is: so as to achieve forms of social coordination and organization in our relevant thought and action. And in part in the pursuit of such organization, practices of accountability will frequently come to the fore.

I will be returning later to related issues about the relation between the underlying social psychology of modest sociality and practices of holding accountable. For now, let

[68] My earlier discussion of the roles characteristic of shared intention appealed to related bargaining but did not appeal, as I do here, to shared deliberation. (See "Shared Intention," in my Faces of Intention at p. 112.) Ideas leading to this addition are in my "Shared Valuing and Frameworks for Practical Reasoning," an essay that is the basis for elements of Chapter 8. Andrea C. Westlund independently emphasizes a contrast between bargaining and what she calls "joint deliberation" in her "Deciding Together," Philosophers' Imprint vol. 9, No. 10 (2009), a discussion from which I have benefitted. However, in discussing "Shared Valuing and Frameworks for Practical Reasoning" Westlund indicates that what I describe there "is a more general phenomenon than what [she has] in mind."(p. 6)

me just note that the roles of shared intention I have highlighted are primarily roles in temporally downstream social functioning, including later shared reasoning and bargaining shaped by these shared intentions. This is a parallel with the planning theory's emphasis on the roles of intentions of individuals in temporally downstream functioning, including later practical reasoning. As in the individual case, we will also want a theory about the kinds of reasoning that can intelligibly issue in such shared intentions.[69] (And, again, in many cases that reasoning will itself be shaped by prior intentions – individual or shared.) But the first step is to say what such shared intentions are. And here I think that, as in the case of the intentions of individuals, it is the roles in temporally downstream functioning that are central. In particular, when we return to our question about what distinguishes shared intentional activity from a corresponding contrast case, our answer will appeal primarily to a structure characterized primarily in terms of certain downstream roles and associated norms.

What norms are associated with these general social roles of shared intentions? Well, it seems plausible that there will be associated norms of social agglomeration and consistency, social coherence and social stability. Roughly, it should be possible to agglomerate relevant intentions into a larger social plan that is consistent, that in a timely way adequately specifies relevant means and preliminary steps, and that is associated with appropriately stable social psychological structures. Failure to satisfy these norms will normally undermine the distinctive coordinating, guiding, and structuring roles of shared intention in our social, practical thought and action. And such a tight connection between conforming to norms and realizing central, characteristic roles is an indicator that there is a characteristic rationality norm at work.

So, as in the case of individual intention, we can expect the social roles characteristic of shared intention to be associated with characteristic norms of – in this case -- social rationality. And we will want to understand both the explanatory and the normative significance of these norms of social rationality.[70]

---

[69] This is the central concern in Natalie Gold and Robert Sugden, "Collective Intentions and Team Agency," Journal of Philosophy 104 (2006): 109-137. I discuss this paper in Chapter 5. And I offer a model of shared deliberation in Chapter 8.

[70] In his "Practical Intersubjectivity," Abraham Roth emphasizes that our sociality involves characteristic norms of interpersonal rationality. It is such norms that Roth means to refer to in his talk of practical intersubjectivity. I am here agreeing with Roth that such norms are central to our understanding of such

There are three complexities here, however. The first concerns the explanatory role of norms of plan rationality. In the case of individual planning agency I supposed that what constitutes the explanatory role of norms of plan rationality is primarily the explanatory role of the at-least-implicit acceptance of those norms by individual planning agents. In the case of these social norms, however, it is not immediately clear how to understand their explanatory role. Should we appeal to the acceptance of related, individualistic norms on the part of those individual participants? to the individual participants' acceptance of these social norms? to the group's acceptance of these social norms? On the view I will be developing all three explanatory modes of social rationality are possible; but the first, individualistic case has an important priority.

Second, which intentions fall under these social norms? Well, at the least, whatever intentions constitute the relevant web of shared intentions. If we are singing the duet together, for example, we will have a shared intention to sing together, and perhaps also shared intentions to sing in a certain key, and in a certain style. But it also may be that each of us has related intentions – say, to emphasize a certain note – that concern only her own contribution to the shared activity. If there is to be relevant coordination, the social norms of agglomeration and consistency need to apply to these intentions, not just the shared intentions, strictly speaking. Further, the social pressure for an adequate specification of relevant means to the shared end can be satisfied in part by intentions that are strictly only about the agent's own contributions to the shared activity. And, finally, the need for stability to achieve the social end will apply broadly to these various intentions. In this way these social norms of agglomeration, consistency, coherence, and stability have a somewhat broad scope.

But not an overly broad scope. The roles of shared intention do not require interpersonal consistency of belief or judgment about the values at stake in the shared activity. The successful coordination of our house painting need not require that our aesthetic beliefs about colors be consistent with each other, though it does require consistency in our plans about which colors to use. It is interpersonal consistency and coherence in plan, not in doctrine, that is central to modest sociality.

---

sociality. Roth, however, argues that we cannot account for these norms within a broadly individualistic model of shared intention. In contrast, one of my main claims, to be defended below, is that we can. See Abraham Seeshu Roth, "Practical Intersubjectivity," in .

This is an aspect of a general feature of modest sociality. Modest sociality involves interpersonal coordination and organization of practical thought and action. But modest sociality is possible in the face of conflict of beliefs or judgments about the right and the good. It is intentions and plans that are at the heart of the coordination and organization characteristic of modest sociality. Or so I conjecture. I will call this the primacy of intention for modest sociality. And insofar as this conjecture of the primacy of intention is theoretically fruitful we have yet a further argument for the theoretical need to include intending in the basic story of agency on which our account of sociality builds.

Turn now to a third complexity concerning the relevant social rationality norms. This complexity concerns the cognitive background with respect to which relevant intentions and plans are evaluated for social consistency and social means-end coherence. We are assuming that the agents are in a position to have knowledge of the minds of others. But other matters must also be in the cognitive background. In particular, these norms on intentions of consistency and coherence apply against a background that concerns, roughly, what is possible and what is effective. And the different participants in a shared intention might have differing views about these matters.

To keep my initial discussion manageable, however, I am going to assume for now, as a simplifying assumption, that the participants have the same beliefs about these matters of possibility and effectiveness, and that it is these beliefs that are in the background when we apply norms of social consistency and coherence on the relevant intentions. Later, in Chapter 8, I will revisit this complexity (as part of an appeal to what I will be calling shared context-relative acceptance).

These complexities in hand, I think we can see that something like these social rationality norms of agglomeration, consistency, coherence, and stability will be associated with the social roles that are characteristic of shared intention. These social norms apply to relevant intentions and plans, given a certain cognitive background.

7. Constructivism about shared intention and modest sociality

So we have structures of roles and associated rationality norms both at the level of individual intention and at the level of shared intention. Our next question is: how are

these structures related? And here, as a first step, it is natural to draw on our third idea: the methodology of creature construction. We try to see the move to these social roles and associated norms as in some way building, within Gricean creature construction, on the roles and norms characteristic of individual planning agency, and in response to pressures for increased coordination and unity of agency at the social level.

This is not to say that in the course of our actual lives we ourselves make a transition from non-social to social creatures. Creature construction is not a story of actual human development, and it can recognize that human lives are embedded in the social from the start -- that, as Pierre Demeulenaere has put it, "the social is always already there."[71] What we are after is not a story of actual human development but an understanding of the conceptual, metaphysical and normative deep structure of our sociality.

This Gricean picture still leaves unsettled, however, how exactly to characterize this step to modest sociality. Recall that the step from individual purposive agency to individual planning agency involves – according to the planning theory – a step to a distinctive nexus of roles and norms. According to the planning theory this step brings with it attitudes of intending that are distinct from forms of wanting and believing characteristic of simple purposive agency -- though no less embeddable in a natural causal order. Now, we have seen that a step to shared intention and shared agency involves an analogous step to a characteristic nexus of social roles and social norms. So we need to ask: to what extent does this step to shared intention and shared agency involve the introduction of phenomena that are fundamentally distinct from those of individual planning agency? To what extent, in contrast, can and should this step build more conservatively on the planning theory of the intentions of individuals?

Well, let us reflect a bit more on the step from

(a) individual desire-belief purposive agency,[72]

to

(b) individual planning agency,

[71] Pierre Demeulenaere, "Where is the Social?" in Chris Mantzavinos, ed. Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice

[72] As in, for example, Creature 3 in my "Valuing and the Will".

This step from (a) to (b) involves the introduction of a form of psychic functioning – namely, planning -- that has an independent impact on thought and action, an impact over and above the ordinary functioning of a simple desire-belief psychic economy.[73] These planning phenomena will, of course, systematically interact with ordinary beliefs and desires, and they need to fit with our basic cognitive limits; but they have, according to the planning theory, their own distinctive roles in the dynamics of practical thought and action.

Now consider the step from (b) to

(c) shared intention and modest sociality.

When we go from (b) to (c) are we moving to a form of social functioning that has an independent impact on thought and action, an impact over and above the functioning of the psychic economies in (b)? Is this move from (b) to (c) analogous in this way to the move from (a) to (b)?

I think that the answer is "no". After all, we do not suppose that shared intention shapes shared action in a way that reaches its hand over the psychic functioning of the individual agents who are involved. We expect that shared intention, whatever it is, works its way through the workings of the individual psychic economies, appropriately interrelated.[74] One way to think about this would be to see shared intention as in some way consisting in relevant, interrelated intentions of the individual participants. And that suggests that we see the step from (b) to (c) as a fairly <u>conservative</u> step in creature construction from individual to shared agency.

Just how conservative this step will turn out to be will depend on how we understand the relevant, inter-related psychic functioning of the individual agents who are involved in shared agency. For now let me just sketch in broad outlines the basic picture of a conservative step from (b) to (c). The idea is to build on structures of individual planning agency primarily by characterizing certain relevant contents of the intentions of the participants, relevant contexts in which those intentions are embedded, and relevant inter-relations among those intentions. The idea, roughly, is that the social-

[73] Such an independent impact is apparently denied by a reductionist account of intention as a complex of desires and beliefs.

[74] This is broadly in the spirit of Philip Pettit's view that "the natural person is the ultimate center of action…" "Groups with Minds of Their Own," in Frederick Schmitt, ed., <u>Socializing Metaphysics: The Nature of Social Reality</u> (Lanham: Rowman and Littlefield, 2003): 167-193, at 190.

norm-assessable social functioning characteristic of shared intention emerges from the individual-norm-assessable and individual-norm-guided functioning of relevant structures of interrelated intentions of the individuals, as those intentions of individuals are understood by the planning theory.

We seek, that is, a construction of inter-connected intentions and other related attitudes of the individuals in appropriate contexts that would, when functioning in the norm-guided ways highlighted by the planning theory of the intentions of individuals, play the roles characteristic of shared intention. And we try to see conformity to central norms of social rationality characteristic of shared intention – norms of social consistency, social agglomeration, social coherence and social stability -- as primarily emerging from guidance by norms of individual plan rationality that apply directly to the relevant interrelated structures at the individual level. If we had such a conservative construction we would have reason to say that this construction is shared intention, or at least one important kind of shared intention.[75] And such a conservative construction of shared agency – if it were available – would pose a challenge to a more top-down approach that begins with the shared case and posits fundamental discontinuities in the step from individual planning agency to shared agency. Such a top-down approach would need to explain the need to appeal to such discontinuities, given the (assumed) success of the more conservative construction.[76] I will return to this last point below.

In describing this approach to shared intention I am implicitly distinguishing between being assessable by a norm, being guided by a norm, being explained by a norm, and conforming to a norm. To think that relevant thought and action is assessable by a norm is to suppose that the norm applies to that thought and action. A norm guides relevant thought and action when its acceptance, or other form of internalization, is an appropriate explanatory aspect of the actual psychological functioning. A standard way – but, as we will see, not the only way -- in which a norm can help indirectly to explain certain psychological functioning to which it applies is by way of such acceptance of that norm by relevant individual agents. Thought and action conform to a norm when the norm applies to them and they do not violate it. In these senses, a norm can apply

[75] I discuss differences between these two ways of putting this point in "I Intend that We J" in Faces of Intention at 144. I return to the possibility of multiple forms of shared intention below in section 8.
[76] In this last sentence I am indebted to conversation with Michael Smith.

without actually guiding or in other ways explain; a norm can both apply and guide even though there is, in a particular case, a breakdown in conformity to the norm; and there can in fact be norm conformity even if the norm does not guide or in other ways explain. My aim is to provide a construction of inter-connected intentions of individuals whose individual-norm-assessable, individual-norm-guided, and individual-norm-conforming functioning (according to the planning theory of individual agency) would constitute and help explain the social-norm-assessable and normally social-norm-conforming social functioning of shared intention.

I have proposed that certain social rationality norms – social norms of agglomeration, consistency, coherence, and stability – apply to shared agency, and that such shared agency will normally conform to those norms. But, as noted earlier, we need to ask: who or what accepts and applies relevant norms? And the answer that is built into the kind of construction I seek is that in the basic case the relevant norm acceptance is that of the individual participants, and the norms accepted are, in the first instance, the norms of individual planning agency. Given appropriate contents, contexts, and inter-relations of the intentions of these individual participants, these phenomena then induce the social-norm-conforming social functioning of shared intention, and the applicability to that functioning of the cited social norms. And when this is the case we can say that the social norm helps explain the social functioning by way of the acceptance of associated norms of individual intention rationality by the participating individual planning agents.

The central idea is not that it is the participants in a shared intention who do this constructing. The participants in a shared intention participate in that shared intention; they need not literally construct it. Nor is the idea that there is an actual historical transition from solely individual planning agency to participation in shared intention. Rather, we the theorists seek to understand what is involved in or constitutes such a shared intention as a structure that consists in certain individualistic elements related in certain ways. We try to do this by constructing this structure of elements. And here we, the theorists, are aided by the Gricean methodology of creature construction pursued in a way that seeks a more or less conservative construction. That said, there will be cases in which it is natural to say as well that the participants themselves intentionally

construct their shared intention by taking steps aimed at creating the conditions that constitute shared intention.

Call the idea that shared intention consists in a structure of relevant and suitably inter-related attitudes of the participants in a suitable context constructivism about shared intention.[77] We begin with the idea that shared intentions interpersonally structure and coordinate thought and action, and that these structuring and coordinating roles involve associated social norms. We then ask: will these social-norm-assessable social roles be grounded in the individual-norm-assessable and individual-norm-guided functioning of appropriate attitudes of the individual participants – attitudes with appropriate contents, in appropriate contexts, and appropriately interrelated? We seek to answer this question by constructing a structure of interrelated intentions of the individuals, and norms that apply to and guide those intentions, that would induce the social-norm-assessable and social-norm-conforming social roles characteristic of shared intention. We want to show that intentions of individuals in these special contexts and with these special and distinctive contents and interrelations would, insofar as they function properly and in a way that is guided by the norms of individual planning agency, play the roles of shared intention in part by issuing in thought and action that conforms to central norms that apply to shared intention. And we want to show that violations of these norms of social rationality will normally be constituted by violations, by one or more participants, of associated rationality norms of individual planning agency.

Such constructivism highlights the idea that the individual participants are assessable and guided by norms of individual planning agency, but that given the special contents of their intentions, and their characteristic interrelations and contexts, this brings with it the applicability of, and (normally) conformity to, corresponding social norms on shared intention. In this specific and limited sense, constructivism posits a kind of normative emergence. When the individuals become aware of this normative emergence they may go on explicitly to internalize these social norms and directly

[77] This is broadly in the spirit of what Scott Shapiro calls the "constructivist strategy" in legal philosophy. See his Legality at 21. A different kind of constructivism is a view about the nature of practical reason. See e.g., Christine Korsgaard, "Realism and Constructivism in Twentieth Century Moral Philosophy," Journal of Philosophical Research 32 (2003): 99-122. Constructivism about shared intention aims to be neutral on these meta-normative issues.

appeal to them in their practical reasoning. And they may do this in part because they can see the advantages that accrue to the group's conformity to those norms, both in general and in the particular case. Their internalization of these social norms would then add a further element to the explanation of conformity to those norms.

Such a step to a second explanatory mode of social rationality would still remain within the domain of the acceptance of norms simply by the participating individuals. As we will see in Chapter 8, however, there is also the possibility of yet a further step, a step to a kind of shared acceptance of these norms of social rationality. Such a shared acceptance of these norms would be a kind of shared policy, in a sense I will explain in that chapter. For now the important point is that all three of these explanatory modes of social rationality are possible, but what is at the bottom of all three is the acceptance by the individual participants of norms of individual intention rationality.

Constructivism does not suppose that all that is important in shared agency is fully grounded in such broadly individualistic structures. Constructivism grants that much of our agency – individual and shared -- takes place within, and is shaped by, larger cultural, political and legal structures. Constructivism grants that there can be distinctive social values at stake in shared agency – for example, the value of certain forms of social unity and social governance.[78] Constructivism grants that shared agency raises distinctive issues of reliance, trust and trustworthiness (as well as issues about ordinary civility). After all, the stability of a shared intention may well depend on the extent to which the participants can reasonably rely on and/or trust each other.[79] And constructivism grants that shared intentions normally activate distinctive moral obligations of each to another, obligations that can lend support to the stability of the shared intention. So the normative emergence posited by constructivism is only a part of the normative story.

What constructivism does say is:

(a) The characteristic functioning of shared intention is in basic cases constituted by the characteristic functioning of relevant structures of

[78] I appeal to some such values in "Dynamics of Sociality". Thanks to Luca Ferrero for noting the need to make this explicit here.

[79] For an approach to shared intention that highlights its connection to reliance see Facundo Alonso, "Shared Intention, Reliance, and Interpersonal Obligations," Ethics 119 (2009): 444-475. .

interrelated intentions of the individual participants in relevant contexts, as that functioning is understood within the planning theory.

(b) The application of central norms of social rationality to shared intention, and the conformity to those norms, emerges in these basic cases from the guidance of the individual participants by the central rationality norms of individual planning agency as those norms apply to the intentions of those individuals, given relevant contents, contexts, and interrelations.

In this way constructivism builds on the planning theory of individual planning agency. It supposes that once we have these distinctive structures of individual planning agency on board, the further step to shared agency can be a conservative step of the sort described. In this sense, while highlighting the great significance of such sociality to our lives, constructivism posits a deep continuity – conceptual, metaphysical, and normative -- between individual planning agency and modest forms of sociality. That, anyway, is the conjecture; the proof, of course, must be in the pudding.

8. Continuity, sufficiency, and Ockham's Razor

I have said that my aims are conceptual, metaphysical and normative. I seek relevant conceptual clarification, articulation, and innovation; I seek a view of the underlying metaphysics of shared agency; and I seek an understanding of relevant normative aspects of shared agency. These three aims have helped shape my discussion so far. But we can now identify a further general aim that lies behind these reflections.

Now we are throughout being guided by an overarching aim of contributing to a fruitful and insightful human social psychology that takes seriously these phenomena of modest sociality. But the project I am sketching here involves approaching that overarching theoretical aim in a specific way. This specific approach to this general theoretical aim is implicit in the pursuit of a conservative construction of shared intention and modest sociality. The idea is to see whether there is an important continuity – conceptual, metaphysical, and normative – between individual planning agency and modest sociality.

This points toward a central concern with relevant sufficient conditions for shared intention and modest sociality. This is because appropriate sufficient conditions would be enough to establish the cited continuity. This pursuit of a conservative construction allows for the possibility of multiple constructions, each of which provides some such sufficient basis for the social roles and norms characteristic of shared intention. What is crucial for this theoretical ambition is to provide at least one such structure of sufficient conditions that satisfies the continuity constraint.

This leaves open the possibility that there may turn out to be multiple such sufficient conditions each of which satisfies the continuity constraint. In the face of such purported, alternative constructions, we would need to ask which makes better sense of the complexities of these forms of sociality. And here we might in the end sensibly give the nod to one of the purported constructions as theoretically more fruitful. But we need to leave open the possibility that the best thing to say might be that shared intention is multiply realizable. This sort of multiple realizability would be compatible with the kind of continuity between individual planning agency and modest sociality that I will be trying to defend.

This willingness to countenance multiple constructions each of which satisfies the continuity constraint does not extend to multiple constructions one of which satisfies the continuity constraint and one of which does not. As anticipated earlier, if we can indeed articulate sufficient conditions for modest sociality that satisfy the continuity constraint, then there will be a presumption in favor of that model of modest sociality in comparison with a proffered model that involves a basic discontinuity.

This is an application of Ockham's Razor. If we can get a plausible model of modest sociality without appealing to a fundamental discontinuity in the step from individual planning agency to such sociality, there is a presumption against an appeal to such a discontinuity in our theorizing. If a conservative construction works then there is a presumption in its favor in comparison with a non-conservative model.

Now, the two main competing views that I will be discussing here – Searle's appeal to an irreducible we-intention, and Gilbert's appeal to an irreducible joint commitment -- are each of them versions of a non-conservative, discontinuity theory. An implication of the cited application of Ockham's Razor is that a successful

conservative construction of modest sociality yields a presumption against the introduction into our theory of such new, irreducible elements. And what is crucial for such a conservative construction is to provide sufficient conditions for modest sociality. Of course, it would also be good to know whether the conditions cited are necessary for modest sociality; but a concern with necessity is less pressing, for reasons noted. For this reason my main focus will be on sufficiency for modest sociality within the constraints of the continuity conjecture.

9. Deception, coercion, shared intentional, shared cooperative

Before proceeding with this project of construction I need to note one more complexity. Sometimes human interactions involve forms of deception or coercion between the participants. And sometimes such deception or coercion blocks the claim that people are acting together in a shared activity.[80] Suppose that you and I are painting the house, but you are deceiving me about central features of how we are proceeding even though I have made it clear that these are features about which I care. For example: though it is important to me that we stay within a certain budget, and I have made it clear that I am acting on the assumption that we are indeed within budget, you know that we are very much over budget. But, in order to keep me engaged in the painting project, you intentionally deceive me about this matter. In a different case, as we proceed I begin to balk and to express a desire to stop the project. Your response is to pull out a gun and coerce me into continuing to paint, even while recognizing that my painting will now be motivated by my fear of your threatened sanction rather than any interest in our joint activity. In each of these cases the deception or coercion between us so infects the inner workings of our interactions that they baffle shared intentional activity. You are instead – as we say – merely using me.

Consider now somewhat different cases of deception and coercion. Suppose we are painting the house together but you lie to me about your reasons for participating. You say you are participating because you dislike the present color; but in fact you are participating to win a side bet. But suppose also that I really do not care what your reasons are for participating, so long as your reasons are not egregious and they do not

[80] My focus here is on deception or coercion between the participants. If some third party got us to sing together by lying to us about the rewards, or threatening harm if we did not sing together, that would not normally prevent our joint singing from being a shared intentional and a shared cooperative activity.

get in the way of your participation. Again, suppose that you are the master, and I a slave. You issue an edict, backed by a threat: we are going to pick the cotton together. In response, I go ahead and work with you in the fields, and we do indeed pick the cotton together. As we proceed the threat remains in the background. I would, of course, much prefer that you not be in such a position of power over me concerning whether we are going to pick the cotton. But I am nevertheless prepared to work with you in our joint activity, and our working together may involve various subtle forms of interaction and adjustment. In each of these cases the deception or coercion does seem to block the idea that we are engaged in a cooperative activity, that we are each cooperating with the other. Nevertheless, it seems possible that the deception or coercion in these cases does not actually interfere with the specific ways in which we interact when we paint, or pick cotton, though of course it might. But it also seems possible that once we start to paint, or pick cotton, the background deception or coercion does not infect the specific ways we actually interact. Though what we are doing seems ill-described as a cooperative activity, it nevertheless seems plausible to describe it as a shared intentional activity.

I do not want to put much weight on linguistic intuitions here. What is important to note is that sometimes deception and coercion between the participants in an activity clearly block the shared-ness of the activity, but that sometimes the matter is more complex. In these more complex cases the deception or coercion between the parties, while it in some ways taints the sharing, need not infect the specific interactions. About these more complex cases I will say that there is shared intentional activity, but not shared cooperative activity. This involves, I grant, a bit of linguistic legislation. But I think it is a plausible way of marking an interesting difference in the relevance of deception or coercion. On this way of thinking, the idea of cooperation brings with it a broad exclusion of deception or coercion between the relevant parties and with respect to the activity, even if that deception or coercion does not infect the specific interactions in ways that block shared intentionality.[81]

[81] Here I diverge from John Searle's thought that "the notion of … collective intentionality, implies the notion of cooperation." "Collective Intentions and Actions," in P. Cohen, J. Morgan, and M.E. Pollack, eds., Intentions in Communication (Cambridge: MIT Press, 1990): 401-415, at 406.

This is a sign that our concept of cooperation is to some extent a moralized notion in the sense that it incorporates certain moral prohibitions on deception and coercion. In contrast, the idea of shared intentional activity that is front and center here is the idea of a distinctive kind of social-psychological organization in our thinking and acting together. Given the somewhat moralized idea of cooperation, it seems likely to turn out that not all shared intentional activities are shared cooperative activities.

That said, the details of our view of the relation between shared intentional and shared cooperative activity must await our overall theory, one that will focus primarily on shared intentional activity. So let us turn to our efforts to develop such a theory by way of a conservative construction of shared intention and modest sociality. Once the basic elements of our theory are on board we can return, in Chapter 4, to these issues about deception, coercion, and cooperation.

Chapter Two: Building Blocks, Part One

Suppose that in a case of modest sociality you and I share an intention to go to New York City (hereafter, NYC) together. What construction of intentions and related attitudes of each would be such that its norm-assessable and norm-guided functioning (as articulated by the planning theory of the intentions of individuals) constitutes the social-norm-assessable and social-norm-conforming functioning of the shared intention? And how does this construction of shared intention enter into a construction of the modest sociality of shared intentional activity? To answer these questions we need to describe the building blocks for this construction. In doing this we will have our eye on forms of social functioning and social rationality that are characteristic of shared intention. And we will be looking for structures at the individual level that will help constitute or generate these forms of social functioning and social rationality.

Before proceeding let me note a complexity. In articulating these building blocks we will in some cases find helpful a certain redundancy: some of the building blocks that we explicitly cite will be implicit in others that we explicitly cite. My strategy will be to allow for such redundancies when they promote a clearer understanding of the construction. For this reason the full statement of that construction will incorporate some such redundancies. Once all the pieces are in hand, however, I will also offer a compressed version of the proposed construction, one that avoids these redundancies.

1. I intend that we J, and circularity

Let's begin by noting briefly two strategies that will not work. We might first try to appeal simply to my intention to go to NYC given that you too so intend, as well as to your intention to go to NYC given that I too so intend. But this structure of coordinated intentions might be simply a case of strategic equilibrium, like walking alongside a stranger.

What if we appeal to the condition that we each judge that our going to NYC is the best option? Or what if instead we appeal to the related idea that we each rank that option highest in our own relevant evaluative ranking? Well, it seems that there could be this package of evaluative judgments or rankings even though neither party has yet decided to act in accordance with it: perhaps one or both of them is still deliberating. Further, this proposal would block shared intentions in cases of disagreement about

what would be best; and it would also block shared intentions that involve weakness of will on the part of one or more of the participants. But it seems that we can share an intention to go to concert C even though you think A is best and I think B is best; and a pair of weak-willed lovers might shared an intention to have an affair even though each thinks this is not the best option. Here, as in the case of individual intention, we need to be careful to avoid an overly simplistic picture of the relation between intention and evaluation.

As anticipated in Chapter 1, what we need instead is the condition that

(i) we each intend that we go to NYC

where the intentions alluded to in (i) are intentions of the sort characterized by the planning theory of the intentions of individuals.[82]

There are two occurrences of "we" in (i). The first is what Christian List and Philip Pettit call "the distributed 'we'".[83] The appeal is to my state of intending and to your state of intending. What about the second occurrence? Well, in basic cases this use of "we" will also be distributive. But we can also, without circularity in our purported construction, here avail ourselves of a concept of a group, more or less precisely specified. We can do this so long as that concept of a group does not itself bring with it the very idea of shared intentionality. I might intend, say, that those of us in this part of the park run toward the hot air balloon that has crashed. So long as this use of "we" (or "us") does not bring with it the very idea of shared intentionality, there need be no vicious circle.[84]

In my understanding of (i) in a model of small-scale modest sociality I will suppose that each of us has the ability accurately to pick out the other participant and identify him as one's partner. I do not merely know of my partner as, say, the richest person in the room whosoever he or she may be. My intention is that we (that is, me and you – where I have the ability accurately to pick you out and identify you as my

[82] In my early thinking about this condition I was helped by remarks of Philip Cohen.
[83] Christian List and Philip Pettit, Group Agency: The Possibility, Design, and Status of Corporate Agents (Oxford University Press, 2011) at 194.
[84] This does not mean that the "we" in these contents cannot ever refer to a group in a yet stronger sense that does bring with it the very idea of shared intentionality. The idea is only that at this basic level of our construction we need to avoid the use of such a stronger sense.

partner) go to NYC.[85] This assumption will make it easier to understand the more complex inter-relations among the participants to which we will be led as we try to enrich the relevant building blocks. Granted, this is an assumption that would need to be relaxed if we were to try to extend the theory to larger groups. But, as I have said, I will be satisfied here if we can agree on a basic approach to small scale cases of modest sociality; and for such cases it is plausible to make this assumption.

Of course, an individual may have an intention he would express as "we will do it," and yet be mistaken that his use of "we" succeeds in referring. But in that case there is no shared intention; so I here put such cases to one side.

A basic point is that appeal to these intentions in (i) ensures that an intention-like commitment to our activity is at work in the practical thinking of each. Each is appropriately settled on and committed to our activity, where we understand such commitment on the part of each in terms of the planning theory. In particular, once our activity is an element in this way in my plans, I will face characteristic problems of means with respect to our activity, given a need for means-end coherence of my plans. I face this characteristic problem about means because I intend our activity. In this way I can be led from my intention concerning our activity to an intention to do something myself as a means to or element in that activity of ours, perhaps as a way of helping you play your role. In contrast, these demands of mean-end coherence would not in general be engaged if each only had some desire in favor of the joint activity. Further, once our activity is an element in my plans I will be constrained by characteristic requirements of plan agglomerativity and consistency with respect to our activity. In this way I can be led to filter out certain intentions on my part to act in certain ways. Again, this is because I intend our activity. In contrast, these demands of agglomerativity and consistency would not in general be engaged if each only had some desire in favor of the joint activity, or even had that activity merely as a goal.[86] It is, then, by appeal in particular to (i) that we can explain something we need to explain, namely: the appropriate, norm-guided responsiveness of the thought and action of each to the end

[85] Thanks to Abraham Roth, who credits Michael Thompson, for encouraging clarification here.
[86] See my discussion of the distinction between plan-states and goal-states in Chapter 1.

of the shared activity, responsiveness that is an element in the characteristic functioning of shared intention.

In appealing to (i) I am making the simplifying assumption that in a shared intention to J the participants will have a common conception of J, that there is in this way a match in what is intended by each. This assumption might be weakened at some point. Perhaps we can share an intention to go to NYC if I intend that we go to NYC and you intend that we go to that very large city to the east, if in fact that very large city is NYC. But these are matters we do not need to settle here.[87]

Might we avoid this appeal in (i) to intending that we act, by appealing instead to each person's intention to act with the other? Could we just appeal, say, to my intention to go to NYC with you, and your intention to go to NYC with me? This is an intriguing suggestion since it seems to bring the other into the content of the intention of each and yet retain the idea that what is intended is, at bottom, one's own action, suitably characterized by way of the with-clause.[88]

Well, what is it to intend to go to NYC with you? When I intend to go to NYC with you, do I simply expect that you will be going and, given that, intend to do my part of what would turn out to be our walking? Or do I intend that you go, as part of our going? If the latter then talk of my intending to go to NYC with you will be fairly close to my talk of intending that we go: in each case my intention extends to your role in our activity. If the former – if in intending to go with you I only expect you to be going and do not intend that you go, as a part of our going -- then I will not thereby have the cited dispositions of thought and action concerning social coherence and consistency, dispositions that are characteristic of shared intention. I will not, for example, thereby be disposed to track means to your going (and so to our going), or to avoid activities

[87] Debra Satz emphasized this issue in conversation. Matthew Noah Smith argues that this need for conceptual convergence induces limits on the applicability of my account of shared intention to larger social phenomena. I am inclined to think that this worry can be to some extent defused by noting that in many cases there is agreement about what we are doing together at a somewhat abstract level even when there is not agreement at a more specific level. But this is a matter for another occasion. See Matthew Noah Smith, "The Law as a Social Practice: Are Shared Activities at the Foundations of Law?" Legal Theory 12 (2006): 265-92.

[88] This is one way to read the appeal to intending to do one's "part" of a joint action in Raimo Tuomela and Kaarlo Miller, "We-Intentions," Philosophical Studies 53 (1988): 367-89. And it is in the spirit of Christopher Kutz's model of "participatory intentions" in his "Acting Together," Philosophy and Phenomenological Research 111 (2000): 1-31.

that are incompatible with your going (and so to our going). So I think we really do no need something along the lines of (i).[89]

This purported dilemma is similar to but importantly different from an objection offered by John Searle to an appeal by Raimo Tuomela and Kaarlo Miller to an intention to do one's part in a joint activity.[90] Searle appeals to an example in which each of many business people pursues her own profit-making activities while knowing that that is what the others are also doing, and expects that (given the "hidden hand") this will all result in overall human happiness. But no one sees herself as cooperating with the others to achieve that overall good. If this is all that is involved in intending to do one's part in bringing about the overall good (as that is understood by Tuomela and Miller) then such an intention to do one's part is too weak to get us to shared intentionality. Searle contrasts this case with one in which each really is cooperating with the others in pursuit of that overall good. Searle supposes that in this second case each does intend to do his part in the joint activity of bringing about the overall good, in the sense that is needed for collectively intentional activity. But Searle thinks that if we appeal to this second reading we will be building the very idea of cooperation and collective intentionality into the content of each person's intention to do his part. Searle concludes that to capture the phenomenon of collective intentionality we need to appeal to an irreducible we-intention.[91]

The dilemma I have posed agrees with Searle that a theory of shared intentionality needs something stronger than the first reading of intending to do one's part in a joint activity; and it agrees with Searle that the reading we need will involve

---

[89] In his "Acting Together" Christopher Kutz provides a model of what he calls a participatory intention, where this is offered as an alternative to -- because weaker than -- my condition (i). Kutz says that to have a participatory intention "participants need not intend to achieve [the] collective end. It is sufficient that participants regard themselves as contributing to a collective end."(21) But if to regard oneself as contributing to an end is not to intend that end but only to believe that what one is doing will in fact contribute to that end, then one will not thereby be set to adjust in response to pressures of consistency and coherence with respect to that end. So one will not thereby be a participant in the kind of shared intention I am trying to understand.

It may be that this difference is due to a difference of philosophical target. Kutz seeks an account of joint action in which certain institutional authorities are the source of much of the social organization ("Acting Together," section VI.), whereas I have put aside cases involving such asymmetric institutional authority relations.

[90] Raimo Tuomela and Kaarlo Miller, "We-Intentions," Philosophical Studies 53 (1988): 367-89; John Searle, "Collective Intentions and Actions,"

[91] John Searle, "Collective Intentions and Actions," at p. 405.

some sort of reference to the joint activity. But Searle's step to the further claim that this appeal to the joint activity must itself involve the very idea of collective intentionality is problematic. As I explain below, the appeal to the joint activity J within the content of an intention can be neutral concerning whether or not J is a shared intentional activity.

Tuomela's 2005 response to Searle's criticism is in a similar spirit. Tuomela indicates that on the view in his essay with Miller, to intend to go to NYC as one's part in one's joint trip to NYC is not just to expect that the other will be going and, given that, intend to go oneself. Instead, Tuomela and Miller seek something like the second understanding of such an intention to do one's part. But Tuomela argues that, pace Searle, we can get this second understanding without an irreducible appeal to the very idea of collective intentionality, and so without an unacceptable circularity.[92]

My proposal is, then, in the spirit of this effort by Tuomela to get a stronger reading of intending to do one's part in the joint action, without an unacceptable circularity. On my view we best do this by appealing to an intention that we J (where, in basic cases, J can be neutral with respect to shared intentionality), and then understanding such intentions in terms of the planning theory.[93] And this thought lies behind condition (i).

But how can such a view avoid an unacceptable circularity? After all, on this approach shared intentional activity will be, in the basic case, activity suitably explainable by shared intention. But (i) is supposed to be an element in a construction of shared intention. So if the concept of our activity that is at work in (i) were the concept of shared intentional activity, there would be a problematic circularity in our construction of shared intention.

It is in light of such concerns about circularity that Searle proposes that we see the concept of shared intentionality as a primitive that enters into the contents of

[92] Raimo Tuomela, "We-Intentions Revisited," Philosophical Studies 125 (2005): 327-369, at note 8 and section VI.

[93] In contrast, Tuomela thinks that "one can only intend to do something oneself in the last analysis" ("What Goals are Joint Goals," at 10); so he seeks an alternative strategy. This alternative strategy, which I will not try to discuss in detail here, is more complex than my proposal. See section VI of this 2005 essay. I agree with Tuomela that we can get a sufficiently stronger sense of intending to do one's part in the joint activity without being subject to Searle's charges of unacceptable circularity; but I am proposing an alternative to Tuomela's specific way of doing this.

relevant intentions of the individuals.[94] But I propose a different tack, one broadly analogous to an approach to a corresponding issue about individual intentional action.[95]

In the individual case we try to understand intentional action primarily in terms of intentions of individuals and the right sorts of connections between those intentions and relevant behavior. But what is the content of these intentions? Well, it is common in our ordinary thought to think of what one intends as something done intentionally. But if the content of the cited intentions must involve appeal to the very idea of intentional action, our approach to individual intentional action seems threatened with circularity.

A first step in responding to this concern is to focus on cases in which we can appeal to a concept of activity that is itself neutral with respect to the intentionality of that activity. Examples include: falling down, shaking, giggling, opening the door, going to NYC, knocking over the table, annoying a friend, scaring those in the room, upsetting the applecart, giving away the secret, harming someone. Concerning such cases we can say, roughly, that one so acts intentionally when one intends so to act (characterized in intentionally neutral terms) and this intention appropriately explains the fact that one does so act.

That said, a person who intentionally gives away the secret may well consciously and explicitly express his intention – to himself and to others -- as an intention to give it away intentionally. He certainly would not say he intends to give it away unintentionally. But given that this person does have an intention or intentions about this matter, we can ask what contents are plausibly assigned to his intentions. And here the idea is that such a person will have underlying dispositions of thought and action that support the attribution to him of an intention whose content draws on a concept of activity that is neutral with respect to intentionality. These underlying dispositions include dispositions

---

[94] Searle claims that "collective intentional behavior is a primitive phenomenon" and we should not seek "a reductive analysis of collective intentionality". ("Collective Intentions and Actions," at 401, 406) And he claims that a fundamental reason for this is that "the notion of … collective intentionality, implies the notion of *cooperation*."(406) His view, I take it, is both that the phenomenon of collective intentional behavior is primitive, and that the "notion" or concept of collective intentionality is not reducible. In contrast, Searle does not take a similar non-reductive tack to individual intentional action. See John Searle, Intentionality (Cambridge: Cambridge University Press, 1983), chap. 3. My approach to collective intentionality is more in the spirit of Searle's approach to individual intentional action than is Searle's own approach to collective intentionality.

[95] I note this parallel in "I Intend that We J," at 147. A worry about circularity in the individual case goes back at least as far as Prichard's 1945 essay "Acting, Willing, Desiring", as reprinted in H.A. Prichard, Moral Writings (Oxford: Oxford University Press, 2002): 272-281, at 274-276.

in thought and action to track and to adjust in support of the cited intentionally neutral activity-type, as well as dispositions of responsiveness and adjustment to conflicts between that intentionally neutral activity-type and others he is similarly set to track. There can be such dispositions of tracking, adjustment, and responsiveness to this intentionally neutral activity-type in the absence of the agent's conscious, explicit conceptualization of what he is doing in terms of this intentionally neutral activity-type. And the idea is that these dispositions support the attribution of the cited sort of intention-content, given the presence of relevant intentions. As we might say, an intention in favor of activity characterized in intentionally neutral terms is at least implicit in his web of intentions about this matter together with these underlying dispositions of thought and action.

But what about cases of intentionally doing things for which there seems not to be a straightforwardly corresponding, intentionally neutral behavioral type: for example praying, or asserting? Here we suppose that an agent with the basic capacity to act intentionally – a capacity we specify in the indicated way – can then go on, typically with the help of a complex culture, to learn new action concepts that are, as it were, intentionally loaded. He can then go on to have intentions whose contents exploit such intentionally loaded action concepts. Nevertheless, at the bottom of these enhanced capacities for intentional agency are capacities for a basic kind of intentional agency we can understand by appeal to intentions in favor of activity characterized in intentionally neutral terms.

Granted, one lesson of the past 50 years of the philosophy of action is that it is difficult to know how to say what counts as an appropriate explanatory relation between intention and action. I am taking it for granted that this relation will be in some sense causal. But what we want is, more specifically, that the intention issues in the action in, as it is said, "the right way". And we do not yet know how to specify what counts as the right way without circularity. I do not try to solve this problem here. If we can solve this problem for the case of individual intentional action we can then go on to see how we should proceed with shared intentional action. But even if we cannot solve this problem for the individual case and must grant, in the end, a kind of conceptual non-

reducibility of individual intentionality of action,[96] we can still go on to see how we should proceed with shared intentional action.

And here the idea is that in at least many cases we have available a concept of our activity that, while it does draw on ideas of individual intentional action, is neutral with respect to shared intentionality.[97] We have, for example, a concept of our walking down the street that involves only the ideas that, roughly, we are each intentionally walking down the street, that our walking is alongside each other and at a comparable pace, and that we are each avoiding collisions with the other. We then use such relevantly neutral concepts in the contents of the intentions involved in our construction of initial cases of shared intention.

In saying this I am assuming that the concept of our activity at work in the contents of relevant intentions in basic cases of modest sociality, (a) is neutral with respect to shared intentionality, and (b) can be articulated using the conceptual resources of the planning theory of individual agency. What is crucial in response to the issue of circularity is (a). And there may be theoretical purposes for which we need concepts of our activity that satisfy (a) but need not satisfy (b).[98] However, given our effort to defend the continuity thesis, we will primarily be interested in locating within the contents of intentions, in basic cases, concepts of our activity that satisfy both (a) and (b). (I will return to this matter in Chapter 7.)

As in the individual case, a participant in modest sociality may consciously and explicitly express her intention in terms of the very idea of shared intentionality. She may say, for example, that her intention favors our walking together, in a sense of walking together that is loaded with respect to shared intentionality. But the idea is that such a person will have underlying dispositions of tracking, adjustment, and responsiveness that, in the context of her relevant intentions, support the attribution to

---

[96] Davidson saw his theory as a view of this kind. As he said, he saw his reductive account of acting with an intention as "not definitional but ontological". See his "Intending," in his Essays on Actions and Events, at p. 88

[97] "Shared Intention," at 114-115; "I Intend that We J" at 146-148. And see Christopher Kutz, Complicity, at 86-88. In "Collectivity and Circularity," Journal of Philosophy 104 (2007): 138-156, Bjorn Petersson interprets my earlier remarks about this idea as appealing to a purely behavioral notion of our activity, a notion that does not even involve the idea that each individual acts intentionally. (See p. 150.) However, my thought is, rather, that we want an idea of our activity that is neutral with respect to shared intentionality but can involve the idea of individual intentionality.

[98] For a view along these lines see Stephen Butterfill, "Joint Action and Development," forthcoming.

her of an intention that draws on a concept of joint activity that is neutral with respect to shared intentionality (and is available to the planning theory of individual agency). These will include dispositions to adjust and compensate in one's thought and action in ways that track that shared-intention-neutral joint activity, and dispositions to be responsive to relevant conflicts with that joint activity. There can be such dispositions of tracking, adjustment, and responsiveness to this shared-intention-neutral activity-type in the absence of the agent's conscious, explicit conceptualization of what she is doing in terms of this activity-type. As we might say, this intention concerning the shared-intention-neutral joint activity is at least implicit in her web of intentions about this matter together with these underlying dispositions of thought and action.

With this intention in favor of the (shared-intention-neutral) joint activity in hand, we then appeal as well to other elements of the construction (elements to be discussed below) to ensure that when these intentions connect up in the right way to the group behavior there is shared intentional activity. As in the individual case, we can then use these initial cases as bases for a permissible conceptual ratcheting to support intentions that involve concepts of shared intentional activity that are not neutral in this way.[99] Examples of such shared-intention-involving concepts of joint activity might include getting married, or playing chess, or dancing a tango. In this way we seek to provide an account of shared intentional action by appealing to the appropriate roles of shared intention, but to explain what shared intention is without using, in the most basic cases, the very idea of shared intentionality in the content of the intentions of each.

I think it is plausible that our construction can in this way appeal to (i) without unacceptable circularity. Or anyway, this is at least as plausible for the case of shared intentionality as is the analogous strategy for avoiding circularity for individual intentionality. Indeed, I suspect it is more plausible. Even if it turns out that the theory of individual intentional action cannot specify the "right way" without appeal to something like the very idea of individual intentional action, we might still succeed in avoiding a

[99] I take it this is in the spirit of what Christopher Kutz describes as "a genealogical account that shows generally how the capacity to engage in collective action emerges out of capacities explicable without reference to collective concepts." See Kutz, Complicity (Cambridge University press, 2000), at 86.

corresponding circularity for the case of shared intentionality.[100] If we were to succeed we would have an important element in a model of shared intentional action that eschews appeal to a purported conceptual primitiveness of shared intentionality.

This response to concerns about circularity does not yet address worries about the violation of the own-action condition. I will turn to these worries in Chapter 3. But first I need to add to our building blocks.

2. Interlocking and reflexive intentions

In shared intention each participant is committed to treating the other participants not merely as elements of the world that need to be taken into account (and who may in turn take into account one's reaction to them), but also as – as it is natural to say -- intentional co-participants in the shared activity.[101] But what is it to treat another as an intentional co-participant? We do not want to leave this as a primitive phenomenon or concept. What then to say?

Well, for me to treat you as an intentional co-participant I need to be able to know about and respond to relevant aspects of your mind. But this is not sufficient. I might be in a position to know about and respond to the mind of a person with whom I interact but with whom I do not engage in modest sociality at all. Think of two soldiers fighting with each other in wartime. Each acts on the basis of his beliefs about the other's intentions and actions, as well as his beliefs about what the other believes about him. And these beliefs about the other's beliefs about oneself can lead to what Thomas Schelling calls "the familiar spiral of reciprocal expectations."[102] But so far neither need be treating the other as an intentional co-participant in a shared activity. A theory of modest sociality needs to understand what else is needed.

Here again it may be tempting to turn to talk of claims or demands of each on the other, of holding each other accountable, of mutual obligations of one to the other, of entitlements of each to the performance of the other. To treat you as an intentional co-participant is to see each of us as having entitlements to make relevant claims or demands on the other to perform, and to see each of us as having associated

[100] Searle's view seems to be the mirror image of this: reducibility at the level of individual intentionality, non-reducibility at the level of shared intentionality.
[101] Cp. Christopher Kutz's appeal to seeing the other as "a partner in a joint enterprise". Complicity at 78.
[102] Thomas Schelling, The Strategy of Conflict (Cambridge: Harvard University Press, 1960, at 87.

obligations to the other. The most familiar home of such claims, demands, accountability, entitlements and obligations is commonsense morality, though Margaret Gilbert has emphasized obligations, entitlements, claims and demands that are not specifically moral.[103]

Now, I fully agree that much of our sociality involves such relations of obligation, entitlement, and associated forms of claiming, demanding, and holding accountable. But the issue here is where these inter-personal normative phenomena best enter into our theorizing about sociality. And my conjecture is that such appeals at this very basic level to obligation and accountability in our theory of modest sociality is overly hasty. If the only conceptual resources we had at the level of individual agency were the resources of the desire-belief model together with common knowledge then it might be difficult to resist the move to mutual obligation, entitlement, and accountability. But the planning theory gives us much more to say prior to such a move. The planning theory provides the conceptual, metaphysical, and normative resources for models of sociality that are stronger than mere desire-belief-common knowledge models, but that do not yet make an essential appeal, at the ground level, to obligation, entitlement, and accountability (though they do, of course, leave room for such important phenomena). And my conjecture is that we get a deeper understanding of our modest sociality by theorizing in detail at this intermediate level before turning to the kinds of obligations and entitlements commonly involved in our sociality.[104]

---

[103] This is a general theme in her work. See for example Margaret Gilbert, A Theory of Political Obligation (Oxford: Oxford University Press, 2006), section 7.4.

[104] This conjecture -- that it is overly hasty to turn here immediately to obligation, entitlement, and accountability – diverges from the approach developed in detail by Margart Gilbert. I will turn to her work below.

My approach is also different in spirit from Stephen Darwall's emphasis on such reciprocal demands, claims, obligations and entitlements. In his The Second-Person Standpoint (Cambridge: Harvard University Press, 2006) Darwall sometimes writes as if he thinks that the move from (i) the "individual's psychic economy" to (ii) the "social," must be a move to such reciprocal demands and the like. (e.g., p. 195) Darwall associates with Hume an approach to (ii) that proceeds primarily in terms of (i). And he associates with Reid an approach to (ii) that proceeds primarily in terms of entitlements to claim, demand, and the like. Darwall does acknowledge that "it seems to be possible for individuals to share intentions .. without" such reciprocal obligations and the like. (201 n. 34) But this acknowledgment has little resonance in Darwall's detailed approach to the social, an approach that is broadly in the spirit of Reid, as Darwall interprets him. In contrast, while I do not deny the importance of these interpersonal normative phenomena, I think that they should not enter immediately at this ground level in our theorizing about modest sociality, and that there is an important theoretical space here between Darwall's Hume and Darwall's Reid.

Consider now a case in which each intends the joint activity – thereby satisfying (i) – and yet it seems that neither treats the other as an intentional co-participant. Suppose that you and I are members of competing gangs and each intends that we go to NYC by throwing the other into the trunk of the car and driving to NYC. Each might assert in, as it were, the mafia sense, that he intends that we go to NYC; and each might say to the other, somewhat ominously, "we are going to NYC". In such a mafia case (i) is satisfied;[105] yet it seems that neither is treating the other as an intentional co-participant.

What's missing? Well, in intending to throw the other into the trunk each intends to bypass the other's intention. Neither intends that the joint activity of their going together to NYC proceed by way of the other's intention in favor of that joint activity. This suggests that one element we need to add to the model of shared intention is the condition that each intends that they go to NYC in part by way of the corresponding intention of the other. This means that the content of the intention of each includes a reference to the role of the intention of the other. In this sense the intentions of each referentially interlock.[106]

So let's add to the model the idea that each intends that the joint activity go in part by way of the relevant intention of the other participant. But what should we mean here in saying that the joint activity goes in part by way of the intention of the other? The idea is that the joint activity both is in accord with and is in part a result of that intention of the other. But in what sense is it a result? Well, one condition is that the joint activity involves the intentional action of the other, intentional action that is guided by that person's intention in favor of the joint activity.[107] But should we also add to the content of the relevant intentions that the way in which the intention of the other in favor of the joint activity helps lead to the joint activity is itself compatible with that joint activity being a shared intentional activity? Well, it is a natural idea that in treating the other as an intentional co-participant what each intends is that the relevant intention of the other

[105] Recall that the idea of our going to NYC, as it appears in (i), is neutral with respect to shared intentionality.
[106] The classic source of ideas broadly in this spirit is H.P. Grice, "Meaning," Philosophical Review 66 (1957): 377-88.
[107] This condition involves the very idea of individual intentional activity; but that is legitimate in the present context.

works its way through to the joint action not in just any old way but, in particular, in a way that is compatible with shared intentionality. But the problem is that if we simply add this as an element of what each intends we seem to be back to worries about circularity. We would be saying that each intends that the other's relevant intention helps lead to the intended joint activity in a way that is compatible with the shared intentionality of that joint activity. So we would be re-introducing the idea of shared intentionality into the content of the intention of each.

In response to these conflicting philosophical pressures I propose a compromise. At some point in the theory we will need to say how the intentions of each in favor of the joint activity need to be connected to the joint action if there is to be shared intentional activity. We are going to need to spell out what the connection condition is. When we do this we will not want simply to say that the intentions of each lead to the joint action in the way involved in shared intentional activity. Instead, we will want to give an informative and non-circular account of the connection condition. So we can now anticipate this later discussion of the connection condition and say that what each intends is that the relevant intention of the other helps to lead to the joint action in a way that coheres with the connection condition, suitably explained. So long as the connection condition is explained without appeal to the very idea of shared intentionality, there is no circle.

There is, however, a more demanding condition on the minds of each.[108] Each intends that the joint action proceed by way of the corresponding intention of the other and in a way that coheres with the connection condition. Granted, such a complex intention need only be, in the sense I have discussed, implicit in the agent's psychology. And it does seem plausible that in a wide range of cases in which there is a relevant intention, there will be relevant dispositions of thought and action that concern appropriate connections between the intention of the other and the shared-intention-neutral joint activity. Nevertheless, we may in the end want to allow for cases of modest sociality in which what each intends is only that the corresponding intention of the other helps lead to the joint action, without appeal in the content of those intentions to the connection condition. But given our primary interest in conservative sufficient conditions

[108] As Martin Stone once emphasized.

for modest sociality, we can reasonably build the stronger condition into the basic model.[109]

These observations support an appeal not only to (i), but also to

> (ii-initial) we each intend the following: that we go to NYC in part by way of the intention of the other that we go to NYC (and that the route from that intention of the other to our joint activity coheres with the connection condition).

There is in this way a semantic interconnection between our intentions in favor of our going to NYC: the content of my intention includes a reference to your intention and to its role in our action. And vice versa. The intentions of each semantically interlock.

This condition should be distinguished from a condition that says that each believes that the other's intention will be appropriately effective. In (ii-initial) it is the content of the intentions of each that includes a reference to the role of the other's intention.

Let me try further to clarify this idea of semantic interlocking by considering an example offered by Seamus Miller. Miller writes:

> Suppose I have as an end that we dig a tunnel under the English Channel. You are in France and will dig from Calais, and I am in England and will dig from Dover. The tunnels will connect in the middle of the Englsh Channel. … I don't care whether you have as an end that we dig the tunnel, or whether you are simply digging a tunnel from Calais to the middle of the English Channel for a bet. …. So I don't have as an end that we dig the tunnel (even in part) because you have as an end that we dig the tunnel, though as it happens you do have as an end that we dig the tunnel. Your sentiments mirror mine.[110]

In the terms of my discussion here, what is suggested here is that each of us intends that we build the tunnel, but neither intends that we build the tunnel by way of the other person's intention that we build the tunnel. After all, neither of us "care[s] whether the other agent has the same joint intention or end, just so long as that agent performs that agent's contributory action."[111]

---

[109] I return in Chapter 4 to related concerns that the model may be over-demanding.
[110] Seamus Miller, Social Action (Cambridge University Press, 2001) at 75.
[111] Seamus Miller, Social Action (Cambridge University Press, 2001) at 75.

But, in fact, in Miller's example it may well be that each intends that they build the tunnel in part by way of the other's intention that they build it. In the example I have no preference as between a case in which you intend that we build the tunnel and a case in which instead you intend only to dig your half in order to win a bet. Still, I know that in fact what you intend is that we build the tunnel. And I might well intend that this very intention of yours play its relevant roles. After all, though I have no preference as between dancing with A and dancing with B, if A is now my dance partner I may well intend that we dance by way of her intentions not by way of B's intentions. Similarly, I might well intend that we build by way of your actual intention that we build, even though I have no preference as between your so intending and instead your intending only to build your half. Semantic interlocking is a relation between our actual intentions, and is compatible with the absence of a preference that favors that intention of the other over other possible intentions that might suffice for achieving the joint activity. So Miller's example need not pose a problem for (ii-initial).[112]

The idea that the content of the intentions of each includes reference to the appropriate efficacy of the intention of the other was part of our effort to capture the idea that, in contrast with the mafia case, each sees the other as an intentional co-participant. But in shared agency each will also see herself as an intentional co-participant. So if each satisfies (ii-initial) with respect to the other, it will be natural to suppose that each will also satisfy an analogous condition with respect to herself. This suggests that the full statement of the condition we want here will be along the lines of

(ii) we each intend the following: that we go to NYC in part by way both of the intention of the other and of this very intention that we go to NYC (and that the route from these intentions to our joint activity satisfies the connection condition).

So, the intentions of each will be both interlocking and reflexive.

(ii) entails our earlier condition (i). This is an instance of the kind of redundancy anticipated earlier. But I think that this redundancy supports a clearer understanding of the relevant building blocks. So I will continue to spell out those building blocks in a

[112] In Chapter 3 I return to Miller's example to see if it poses a problem for the account of interdependence developed there.

way that retains this redundancy, though – as promised earlier – once all the pieces are in place I will also provide a compressed version of the construction, one that avoids such redundancy.

Now, a number of philosophers have thought that intentions are quite generally reflexive: intending to A is, quite generally, intending to A by way of this very attitude.[113] The present idea that reflexivity is part of a conservative construction of shared intention need not involve this idea that intentions are quite generally reflexive. The pressure for reflexivity specifically in the shared case comes from the need for inter-personal interlocking in the shared case, plus the apparent symmetry of attitude toward the other and toward oneself. And this pressure is not present quite generally for all cases of intention. So the appeal to condition (ii) does not require (though it does not preclude) the very general view that all intentions are reflexive.

3. Intended mesh

In cases of individual or shared intention the agent or agents will normally have, or be on their way to adopting, relevant sub-plans. For example, perhaps when we intend to paint the house together I have a sub-plan of bringing the paint, and you have a sub-plan of bringing the ladder. I now want to reflect on the attitudes of each toward such sub-plans of the different participants in shared intention.

Return to our shared intention to go to NYC together. There can be cases in which each of us intends that we go to NYC in part by way of the corresponding intention of the other, and yet one or both of us intends to side-step or override, perhaps using deception or coercion, the sub-plans of the other. Perhaps I intend that we go in part by way of your intention that we go, but I intend to trick you into taking the train despite your firm intention to drive. Since I intend to by-pass your sub-plan, I do not intend that we go to NYC by way of sub-plans that are jointly compatible. But it seems that in shared intention there will be, in contrast, a tendency to track and to conform to a norm of compatibility across the relevant sub-plans of each. This is tied both to the

[113] Gilbert Harman, "Practical Reasoning"; Gilbert Harman, Change in View (MIT Press, 1986), chap. 8; Gilbert Harman, "Desired Desires"; Alan Donagan, Choice: The Essential Element in Human Action (London: Routledge & Kegan Paul, 1987), 88; John Searle, Intentionality (Cambridge University Press, 1983); J. David Velleman, Practical Reflection (Princeton University Press, 1989); Abraham Seshu Roth, "The Self-Referentiality of Intentions," Philosophical Studies 97 (2000): 11-52; Kieran Setiya, Reasons Without Rationalism (2007) Part One. For an important critique see Alfred Mele, Springs of Action: Understanding Intentional Behavior (Oxford University Press, 1992), chap. 11.

coordinating role of shared intention and to the mutual treatment of each other as an intentional co-participant.

We can express the point by appeal to a standard form of functioning of shared intention. If we share an intention to go to NYC, and if you intend that we go to NYC by driving while I intend that we go by train, we have a problem. We will normally try to resolve that problem by making adjustments in one or both of these sub-plans, perhaps by way of bargaining, in the direction of co-possibility. So we want our construction to account for this standard norm-conforming functioning of the shared intention.

A first step is to introduce the idea of sub-plans that mesh. The sub-plans of the various participants mesh when it is possible that all of these sub-plans taken together be successfully executed. We can then use in the construction the idea that each participant not only intends the shared activity, but each also intends that this shared activity proceed by way of meshing sub-plans of those intentions of those participants. We appeal, that is, to the condition that

(iii) we each intend the following: that we go to NYC by way of each of our intentions in favor of going to NYC and sub-plans of each of those intentions that mesh with each other.

In this way our construction can ensure that each is committed to, and so appropriately responsive to, the consistent, coherent and effective interweaving of the planning agency of one another in a way that tracks the intended joint action.

Yours and my sub-plans can mesh even if they do not match. Perhaps your sub-plan specifies that we not go during rush hour, whereas mine leaves that issue open; yet our sub-plans are co-realizable. Further, what is central to shared intention is that we intend that we proceed by way of sub-plans that mesh. This can be true even if, as we know, our sub-plans do not now mesh, so long as we each intend that in the end our activity proceed by way of a solution to this problem.[114] Nor need we each be willing to accept just any specification of activities of each that would suffice for the intended end. There may well be, for each of us, ways of achieving the intended end that are unacceptable, and this may manifest itself in conditions that are at least implicit in the

[114] Which is not to say that in every case each must intend that we, in a shared intentional way, solve this problem: that way lies regress (as Jeffrey King once emphasized).

sub-plans of each. If some such condition is violated by the sub-plans of the other then there is a breakdown in mesh.

A further point is that these intentions in favor of mesh do not need to specify precisely how this mesh is to be achieved. Sometimes we achieve mesh by way of our common understanding of what certain types of activity involve – what it is, say, to dance a tango rather than a polka. This can bring to bear various culture-specific conceptions. And sometime we achieve mesh in part because of the way in which some object in the world is responding to our efforts – for example, how the piano is moving as we go up the stairs with it.[115] That said, we may also sometimes find ourselves in conflict about relevant sub-plans. In small-scale cases of such conflict we would normally negotiate or bargain in some way, and our commitment to mesh will be in the background of such negotiation or bargaining. But we might resort to other – as Scott Shapiro calls them – "mesh-creating mechanisms".[116] Perhaps we agree to binding arbitration.

These intentions of each in favor of inter-personal mesh in sub-plans are, as it were, anchored in the intentions of each in favor of the relevant joint activity – in our example, the joint activity of going together to NYC. It is sub-plans with respect to this intended end that are to mesh. If a participant were to give up that intended end – give up, in our example, the end of the joint traveling to NYC—there would no longer be a commitment to relevant mesh in sub-plans (unless there were some other, relevant, more abstract joint action – say, going somewhere or other together – that remained intended by each.)[117]

Does (iii) add a substantive condition that goes beyond (ii), or does it only make explicit what is already implicit in (ii)?[118] Suppose, as in (ii), that I intend that we act by way of your intention that we act and my intention that we act, and by way of a route from our minds to our joint activity that satisfies the connection condition. Do I thereby intend that our joint activity proceed by way of sub-plans that mesh?

---

[115] In these last three sentences I have benefitted from discussion and correspondence with John Campbell, Victor Caston, and Webb Keane.
[116] Scott Shapiro, "Law, Plans, and Practical Reason," 428. Shapiro argues that a central role of authority is that of a mesh-creating mechanism.
[117] For a different approach to related matters see Abraham Roth, "Practical Subjectivity," at 78-9.
[118] A question from Abraham Roth.

Our answer will depend on the specific account we give of the connection condition, as that is included in the content of the intentions in (ii). This is a matter to which I turn in Chapter 3. What we will see there is that given my proposed account of the connection condition, condition (iii) is indeed implicit in condition (ii), suitably understood. This will turn out to be, then, another case of the kind of redundancy anticipated earlier. However, as with the redundancy between (i) and (ii), it will nevertheless promote understanding to keep (iii) as an explicit element of the construction while also providing a compressed construction that avoids this redundancy.

What if I plan to achieve mesh in our sub-plans by coercing you to proceed in a certain way? Here is an example from Facundo Alonso: We each intend that we go to NYC together by way of sub-plans that mesh. You begin by intending that we go by car; I begin by intending that we go by bus. I then threaten you that unless we go by bus I will destroy your reputation; and you acquiesce. Our resulting sub-plans both specify that we go by bus. Do our sub-plans mesh?

I think that in most cases the answer will be No. This is because the sub-plans of most planning agents will at least implicitly include a non-coercion condition. This is a special case of the general point, noted earlier, that the sub-plans of each may include conditions on the acceptability of ways of achieving the end. So if I insist on coercing you in this way, I evidence the absence of an intention that we proceed by way of sub-plans that mesh.

Finally, consider competitive activities.[119] We might be engaged in a shared intentional activity of playing chess together, even though – since we are in competition – neither intends that there be mesh of sub-plans all the way down. This limits the extent to which what we do together is a cooperative activity; but it does not block a shared intention to play chess together. So there will be shared intentions that involve intentions on the part of each that only favor mesh in sub-plans down to a certain level. Nevertheless, given our interest in sufficient conditions for modest sociality, I will focus on cases that involve intention-like commitments to mesh all the way down.

4. Intending, expecting, and the disposition to help

[119] See Searle, "Collective Intentions and Actions," at 413-414.

Suppose that I intend that we go to NYC in part by way of your intention that we so act and meshing sub-plans. My intention engages norms of means-end coherence and consistency. This puts rational pressure on me both to track necessary means to this intended end and to filter further intentions accordingly.

This is so far just to apply the planning theory of the intentions of individuals to the intentions cited in (i)-(iii). But now we need to address a further issue. As I know, our going together to NYC involves actions both of mine and of yours, in each case actions that are explainable in part by my intentions and your intentions respectively. So we can ask: when I intend that we go to NYC, do I thereby intend both that I go (by way of my intentions) and that you go (by way of your intentions)?

Well, sometimes we intend something that involves a certain pre-condition but do not intend that pre-condition: I might intend to respond to your threat or to your offer, but not intend your threat or your offer. Your threat or offer is only a pre-condition of what I intend, not itself something I intend.

However, as I am understanding it, my intention that we go to NYC does not see your contribution to our joint activity as merely an expected pre-condition of our going, a pre-condition to which I am, as Nicholas Bardsley puts it, simply "adding-on" and "providing the finishing touch".[120] Your contribution to our going to NYC is, rather, literally a part of what I intend. In intending, in a way that satisfies my side of (i)-(iii), that we go to NYC, part of what I intend is that we both go, where that involves your going in part by way of your intention that we go.[121]

This means that the demands of means-end coherence and of consistency apply as well to my intention in favor of, inter alia, your playing your role in our joint activity: I am under rational pressure in favor of necessary means to that, and in favor of filtering options incompatible with that. I am under rational pressure in the direction of steps needed as means if you are to play your role in our joint activity. And I am under rational pressure not to take steps that would thwart your playing your role. This will normally

[120] Nicholas Bardsley, "On Collective Intentions: Collective Action in Economics and Philosophy," Synthese (2007): 141-159, at 145. In his positive view Bardsley rejects the idea, in my next sentence, that what I intend really does include your action. See his p. 152.

[121] Can I really intend that? The worry here is, again, the violation of the own-action condition. I return to this issue in Chapter 3.

mean that, insofar as I am rational, I will be to some extent disposed to help you play your role in our going to NYC if my help were to be needed.

A qualification, however, is that I can intend our going, and so your role in our going, and still be willing to bear only a very small cost in helping you; though, if these limits on my willingness to help were public this might make our shared intention rather unstable. And this raises the question: Suppose I am unwilling to incur any costs at all in helping you if need be, though I remain fully confident that you will not need my help and so will in fact play your role without my help? Would such a complete unwillingness on my part to help you, if need be, be compatible with my intending (and not just expecting) our going to NYC, and so your role in that, given that I fully expect that you will not need my help?

Well, I think we can make sense of some such case of intending -- but just barely. We would need to insist at least that, insofar as I am rational, I am set to give up my intention that we go to NYC (and that you play your role) if I were newly to come to believe that you do indeed need my help. Further, if we are to make good on the idea that I intend that you play your role, and do not just expect that you will, we would need to add that at least I am set to filter out options incompatible with your playing your role. So, at the least, I am set not to thwart you in your role. But the idea of being set not to thwart you and yet not being set to help you if need be, is at best tricky. After all, if you need my help then I thwart you if I do not help. In this respect the filtering role and the means-tracking role of intending tend to go together.

So such cases of my intending our joint action (and your role in it) and yet having no disposition at all to help you if need be, are, if possible at all, precious and unusual. An intention that satisfied one participant's side of (i)-(iii) but involved no such disposition to help if needed would be, at most, an attenuated intention. So, given our concern with robust sufficient conditions for shared intention, I will proceed by assuming that the intentions of the participants that satisfy (i)-(iii) are of the ordinary, non-attenuated sort. These intentions involve, by way of rational pressures toward means-end coherence and consistency, at least a minimal disposition to track and to intend means to the joint action, including means to the other's contribution to that joint action. For this reason these intentions involve at least a minimal (though perhaps only

minimal) disposition to help if needed.[122] And this is an aspect of our construction that we will want, since in the absence of any disposition to help if need be, it is not clear that I would be seeing you as an intentional co-participant in our joint activity.

5. Out in the open

Analogues of (i)-(iii) will be basic building blocks in our construction. Given the planning theory, these intentions of each will help ensure modes of norm-guided functioning that are characteristic of shared intention. These modes of functioning will include intention-like responsiveness of each to the end of the shared action, the pursuit of coherent and effective interweaving of sub-plans, and at least minimal dispositions to help.

The next point is that in shared intention the fact of the shared intention will normally be out in the open: there will be a public epistemic access to the fact of shared intention. Such public epistemic access to the shared intention will normally be involved in further thought that is characteristic of shared intention, as when we plan together how to carry out our shared intention. Since such shared planning about how to carry out our shared intention is part of the normal functioning of that shared intention, we need an element in our construction of shared intention whose functioning involves some such thinking of each about our shared intention.

It is here that something like a common knowledge condition seems apt, at least given our primary interest in sufficient conditions. As noted in Chapter 1, there are different approaches to common knowledge, and my hope is provide a theoretical framework for thinking about modest sociality that is available to many such approaches. To fix ideas, however, we can here think of common knowledge as consisting in a hierarchy of cognitive aspects of the relevant individuals: it is common knowledge among A and B that p just when (a) A knows that p, (b) B knows that p, (c) A knows that B knows that p, (d) B knows that A knows that p, (e) A is in an epistemic position to know that (d), (f) B is in an epistemic position to know that (c), and so on. And what we want is that a constituent of our shared intention to J is a form of such

[122] This represents a slight revision of my discussion of the case of the unhelpful singers in my "Shared Cooperative Activity" in Faces at 103-105. In the terms of that earlier discussion, I am now building it into the (non-attenuated) intentions that satisfy (i)-(iii) that there is at least minimal cooperative stability. In my thinking about this adjustment I was helped by discussions with Facundo Alonso.

common knowledge of that very intention.[123] But we do not want to re-introduce problems about circularity by explicitly including in the content of the individual attitudes that are involved in the common knowledge the very idea of shared intention: we do not want simply to say, for example, that each knows that they share the intention. This suggests that we appeal to common knowledge whose content is, more precisely, that the cited multiple components of the shared intention are in place. And we can do this by adding as a further building block:

(vii)[124] there is common knowledge among the participants of the conditions cited in this construction.

It is important that such knowledge is primarily about intention and belief, and not about each person's "vasty deep".[125] If shared agency were generally to require a depth-psychological knowledge of the minds of the participants, shared agency would be much more difficult and much less common than it is. But knowledge of relevant intentions and beliefs seems less problematic and more within ordinary human cognitive limits.

On the assumption that the required common knowledge involves the relevant knowledge of each participant, condition (vii) induces a tight connection between shared intention and each participant's knowledge that the conditions obtain that constitute the shared intention. How does this connection compare to the connection between intention and knowledge in the case of an individual agent? Suppose that I intend to go shopping on Tuesday. Normally, I will know that I so intend. And my knowledge will have two features. First, it will normally not be based on the kind of evidence I usually need to arrive at knowledge of another person's intentions. I normally have some sort of special standing with respect to, or epistemic access to, my own intentions – though exactly how to fill in this idea is a matter of controversy. Further, when I think "I intend

[123] Note that my argument for this has been an appeal to the downstream functioning of a shared intention. Margaret Gilbert also appeals to a common knowledge condition for "joint commitment," and so for shared intention. But her idea is rather that common knowledge is needed in the normal etiology of a joint commitment. See her "What Is It for Us to Intend?" at p. 21. (Facundo Alonso also notes this contrast in his "Shared Intention, Reliance, and Interpersonal Obligations" at note 45 .)

[124] I skip to (vii) to leave room for further conditions, to be discussed below.

[125] See Harry Frankfurt, "The Faintest Passion" at p. 101 quoting from Shakespeare, Henry IV Part I.

to go shopping on Tuesday" I normally seem to be, as Sydney Shoemaker puts it, "immune to error through misidentification" of whose intention is at issue.[126]

I do not say that whenever I intend A I know that I so intend. It seems possible, for example, to misidentify the intended shopping day: while I think I intend to shop on Tuesday I am being absent-minded and what I actually intend is to shop on Monday.[127] Nevertheless, I will normally know what I intend, and my knowledge will normally have the cited pair of features.

Now, in the case of shared intention, my knowledge of the conditions that constitute our shared intention will involve my knowledge about relevant intentions of the others. And this knowledge will be importantly different from my normal knowledge of my own intention. My belief that you intend that we J, if it is to be justified, will normally need to draw on standard sorts of evidence. And there is here no immunity to errors of misidentification of the others. So my knowledge of the conditions that constitute our shared intention will normally draw on ordinary sources of evidence, and is in a context in which "error through misidentification" remains possible. In these two important respects, the participant's knowledge of the conditions that constitute a shared intention, knowledge that is itself a constituent of that shared intention, differs from the kind of knowledge an individual normally has of his own intentions.

Conditions (i)-(iii) and (vii) provide basic building blocks for our constructivism. But now we need to return, as promised, to our rejection of the own-action condition. And this will lead us to reflect on important forms of interdependence between the intentions of the participants in a shared intention.

[126] Sydney Shoemaker, "Self-reference and self-awareness," as reprinted in his Identity, Cause, and Mind (Cambridge: Cambridge University Press, 1984) at 7.

[127] I explore some implications of this point in my "Intention, Belief, Practical, Theoretical," in Jens Timmerman, John Skorupski, and Simon Robertson, eds. Spheres of Reason (2009).

Chapter Three: Building Blocks, Part Two

1. I intend that we J, and the own-action condition

Conditions (i)-(iii) appeal to intentions that violate the own-action condition. Does this ground an objection to these conditions? Well, what might lead a theorist to accept the own-action condition on intention? Granted, if we use only the infinitive construction – intending to – then it will seem that we are indeed limited in this way. But we also have the idea of intending that.[128] We are, for example, at home with talk of a parent's intention that his son clean up his room, or a teacher's intention that the class discussion have a certain character. We are at home with the idea that my intention can sometimes concern and more or less ensure a complex that involves the intentional activity of others. And it is no defense of the own-action condition simply to say that in intending X one believes or supposes one's intention will (likely) lead to or ensure X: I can believe my intention that my son clean up his room will lead to his cleaning the room. If there are, nevertheless, persuasive grounds for the own-action condition on intending, they will require some further articulation and development. So let us try to understand what such a basis for the own-action condition on intending might be. We can then try to respond to these concerns.

Without claiming to exhaust the field, I will first discuss two primary sources of the apparent plausibility of the own-action condition on intention. I will then turn to a somewhat different but closely related objection to my appeal to condition (i).

The first idea comes from Frederick Stoutland, who writes:

> An agent cannot intend to A if she is not prepared to take full responsibility for having done A intentionally …
>
> This condition can be met only if the agent who intends the action is the agent whose action fulfills it.[129]

I take it that Stoutland aims to be offering an argument for the own-action condition that does not just presuppose that all intending is intending to – an argument that appeals to a connection between intending and taking full responsibility. I am, however, skeptical about both elements in this argument. Begin with the idea that to intend X I need to be

[128] See Bruce Vermazen, "Objects of Intention," Philosophical Studies 70 (1993): 85-128.
[129] Frederick Stoutland, "Critical Notice of Faces of Intention" Philosophy and Phenomenological Research at 241.

prepared to "take full responsibility" for X. Now, I agree that adult humans normally should be prepared to take full responsibility for an intended X, in the sense of being prepared to be held accountable for X. Further, others normally take a person's speech act of expressing her intention as a way of taking responsibility. But I do not see this as essential to intending, though it may well be essential to the normal social functioning of adult human beings who are intenders. Intending, as I see it, is to be understood in terms of its roles in planning agency. Its role in practices of accountability seems to me, while obviously of the first importance, not essential to what intending is -- though such practices of accountability would make little sense if intending did not play something like the roles highlighted here. It seems to me that you can be a planning agent while being quite reticent to take responsibility for what you plan to do: someone can be a planning agent and yet refuse to treat himself as accountable at all. Perhaps a certain kind of sociopath is like this. Such a person would be cut off from important aspects of the social world; but such a person might still be a planning agent. To deny that such a sociopath could be a planning agent would be to moralize[130] the very idea of intending in ways that seem to me not plausible.

This is a version of a general issue that we touched on earlier in our initial reflections of the defining roles of shared intention. Both for individual and for shared intending there may be a temptation to see a connection to accountability as an essential feature. Stoutland's idea here seems to be a version of such a view. But I think that we get a better understanding of these most basic elements in our psychic economy by seeing the connection to accountability as grounded in more basic roles in thought and action, rather than as definitive of the very phenomenon of intending.

I do not recommend this strategy for all talk of intentionality, by the way. In earlier work I urged, drawing on work of Gilbert Harman, that our commonsense idea of acting intentionally is indeed tied to judgments about accountability. I saw this as one reason to be wary of an overly tight connection between talk of intending and talk of intentional action, and so to be wary of what I called the "simple view".[131] That is why I talked of "two faces of intention", one – as in the case of the verb 'to intend' -- tied to

[130] Or anyway to quasi-moralize, since there may be non-moral forms of accountability.
[131] See my "Two Faces of Intention," Philosophical Review (1984), and my Intention, Plans, and Practical Reason at 124-5. For recent examinations of related ideas see Joshua Knobe…

psychological explanation and understanding, and one – as in the case of the adverb "intentionally" -- to some extent responsive to concerns with accountability. But here my focus is specifically on intending and its roles in practical thinking, roles that help explain our agency, both individual and shared.

In any case, I am skeptical also of Stoutland's second idea that I can be prepared to take full responsibility for X only if X is my action. Perhaps I will be prepared to take sole responsibility for X only if X is my action. But I do not see why I can be prepared to take full responsibility for X only if X is my action. If you and I conspire to rob a bank together I can be prepared to take full responsibility for this joint robbery, and so can you; though neither of us should think he is solely responsible.[132]

So I think that appeals to the relation between intending and taking responsibility should not lead us to endorse the own-action condition on intending. Granted, in intending X I do see myself as, if all goes well, a kind of causal source of X, whether or not I am prepared to be held accountable for X. In this sense I see myself as, if all goes well, causally responsible for X (which is not to say solely causally responsible). This raises the question of whether I can coherently see myself as, in a relevant sense, causally responsible for our joint action while you too, in our shared agency, also see yourself in that way.

The answer to this last question is, I think, is "yes". But before defending that answer let me consider a second potential ground for the own-action condition on intention. This potential ground derives from the thought that intending involves an anticipation of the experience of acting. We can express the thought this way: when I intend to A I am normally in a position to "anticipate experiencing,"[133] from the perspective of he who is acting, the performance of A. In violating the own action

[132] Harry Frankfurt makes a closely related point in his "Freedom of the Will and the Concept of a Person," at p. 25 note 10.

[133] The expression "anticipate experiencing" comes from J. David Velleman, "Self to Self," first published in 1996, as reprinted in his Self to Self: Selected Essays (Cambridge University Press, 2006): 170-202, at 194. In this discussion Velleman articulates something like the worry I am trying to address here. He says that "framing an intention … entails representing the intended action from the point of view of the agent who is to perform it." (Though note that I appeal not simply to "the point of view of the agent" but to the experience of acting from the perspective of the agent.) And he goes on to say that "the agent who is to perform any action that I intend must be me, since I can't intend the actions of others."(196) But he gives up on this in his later "How to Share an Intention," (first published in 1997 and reprinted in his The Possibility of Practical Reason, at 205), and then in his review of my Faces of Intention in The Philosophical Quarterly (2001): 119-121.

condition a purported intention that we J fails to involve this connection between intending X and being in a position to anticipate experiencing, from the perspective of he who is acting, the performance of X. This is because I am not in a position to anticipate experiencing, from the perspective of he who is acting, our performance of J. And this is because there is not something which is a group's experience of acting. I am not in a position to anticipate experiencing, from the perspective of us, our performance of J, since there is no such thing as the group's experience of acting. But – or so it is averred -- this connection to anticipating experiencing, from the perspective of he who is acting, the execution of an intention, is a basic feature of intention. So it is a mistake to appeal to my intending that we J. Or so it is claimed.

I think, however, that this objection over-generalizes. It may be that to intend to A I need to be in a position to anticipate experiencing, from the perspective of he who is acting, the performance of A. But this is not essential to intending that -- as when I intend that my son clean up his room, or the mafia member intends that the other person go to NYC. We can still see intending-that as intending, though of course differences in content between intending-that and intending-to will be important. We can see intending-that as intending because we can expect that intending-that will be suitably embedded in the relevant, planning-theoretic nexus of roles and norms. Intending-that will respond to a demand for means-end coherence by posing problems for means-end reasoning; it will respond to demands for agglomeration and consistency by filtering incompatible options; it will normally involve associated tracking and associated guidance of thought and action; it will be subject to norms of stability; and so on. And it is by virtue of being embedded in this nexus of roles and norms that an attitude qualifies as one of intending – in contrast with, for example, ordinary desire or ordinary belief. It is intending that we J – not intending-to -- that is needed in condition (i) of our theory of shared intention; and in intending that we J I need not anticipate our experience of J-ing.

I now want to turn to a purported objection to my appeal to (i) that does not insist that all intentions satisfy the own-action condition. What it insists is, rather, that intentions do need to satisfy a closely related condition, one we can call the "settle

condition".[134] And the objection is that it is not clear how, when we share an intention to J, both of our intentions that we J can satisfy this settle condition.

To get at this purported objection let's suppose, for the sake of argument, that my intending X settles that X both in my plans and, thereby, in my beliefs. My intending X settles that X in my beliefs in the sense that I believe that because of my intention it will be that X.[135] I should say that I do not think this is quite right: there seem to me to be cases of intending X in which one does not believe that the intention will be successful, and so it is not in the relevant sense settled in one's beliefs that one will X.[136] But let us put these qualms aside for now and suppose there really is this condition on intention. This is the settle condition.

Applied to the intentions cited in (i) this settle condition entails that my intention that we J settles whether we will J both in my plans and in my beliefs; so, in particular, I believe that my intention will lead to our J-ing. Similarly, if you are my partner in our shared intention, your intention that we J settles whether we will J both in your plans and in your beliefs; so, in particular, you believe that your intention will lead to our J-ing. Further, in a standard case of shared intention we will both be right in these suppositions. But how could we both be right?[137]

---

[134] For this terminology see my "I Intend that We J," where I am responding to a challenge by David Velleman that the intentions appealed to within my theory do not satisfy this condition.

[135] Gilbert Harman thinks that the need for a belief that success depends on the intention applies only to what he calls "positive" intention. Harman notes that some (non-conditional) intentions are not quite like this, and introduces the label "negative intention" for such cases. (See Gilbert Harman, Change in View 80-82.) But to keep the present discussion manageable I will suppose that the intentions of interest here will be "positive" intentions in, roughly, Harman's sense.

Note that it is not being required here that in intending X one believe that X only if one so intends. Though such a belief is common, taken strictly it does not seem necessary. I might believe that even if I do not intend to wear down my sneakers I will nevertheless wear them down because of my intention to run the race. Though this is what I believe, I can still intend to wear them down.

[136] See Intention, Plans, and Practical Reason pp.

[137] As J. David Velleman forcefully puts the worry: "how can I continue to regard the matter as being partly up to you, if I have already decided that we really are going to act?" See J. David Velleman, "How to Share an Intention," as reprinted in his The Possibility of Practical Reason at 205. Velleman acknowledges in this paper, however, that my reply to this challenge in my essay "I Intend that We J" – a reply I am more or less recounting here (with some small changes) - is successful. See his note 11; and see his review of my Faces of Intention in The Philosophical Quarterly.

In his "On Collective Intentions: Collective Action in Economics and Philosophy," Nicholas Bardsley simply assumes that the own-action condition is obvious and "uncontroversial" (144) and he cites Velleman's discussion in "How to Share an Intention," as support for this assumption. (150-151) (Bardsley does not note the distinction between the own action condition and the settle condition.) But, as I have just noted, in his footnote 11 Velleman withdraws this support; and in any case my argument is that it is a mistake simply to assume without argument that this own-action condition is "uncontroversial".

My answer has two stages. In the first stage[138] we appeal to two main ideas. The first is the idea that if we each were to intend J then J would indeed ensue in an appropriate way. And the second is the idea that the intentions of each in favor of J depend on each other in an appropriate way. Roughly speaking, each would rationally adjust his intention in favor of J in response to his recognition that the other has adjusted his intention that we J: in this sense, the persistence of each agent's intention that we J depends causally and rationally on that agent's knowledge that the other also continues so to intend.

So consider a situation in which:

(a) we each intend that we J,

(b) for each of us the persistence of his own intention that we J depends causally and rationally on his own continued knowledge that the other also so intends, and

(c) if we do both so intend, then as a result (and in accordance with the connection condition) we will J.

In such a situation my intention that we J (in (a)) causally and rationally supports the persistence of your corresponding intention (by way of the dependence cited in (b)), and these intentions of each of us in favor of our J-ing lead to our J-ing (as in (c)). When my intention that we J would lead in this anticipated way to our J-ing, in part by way of its support of your intention that we J, the control my intention has on our J-ing (where our J-ing includes both my contribution to our J-ing and your contribution to our J-ing) goes in part by way of your intention that we J. Since the continued presence of your intention that we J is dependent on my intention that we J, we can say that my intention settles our J-ing in part by way of your other-agent conditional contribution to our J-ing. And we can say something similar about the way your intention that we J settles our J-ing in part by way of my other-agent conditional contribution to our J-ing[139]

[138] The second stage is in the next section.

[139] I do not say that we each only have the conditional intention that we J if the other so intends. On the model I am proposing each (non-conditionally) intends that we J, but the persistence of this intention is dependent on the other's so intending. For an appeal instead to corresponding conditional intentions see David Velleman, "How to Share an Intention," at 209 and 215. For a critical discussion of an appeal here to conditional intentions see Abraham Sesshu Roth, "Shared Agency and Contralateral Commitments," Philosophical Review.

Suppose then that (a)-(c) are true and out in the open amongst us. Each of us is in a position to know that his intention that we J will lead to our J-ing, where this will happen in part because of an other-agent conditional contribution. But the role of such a known other-agent conditional contribution in the route from the intention of each to our activity of J-ing need not baffle each person's intention that we J. In knowing (a)-(c) each knows that his intention that we J will appropriately lead to our J-ing in part by way of its support of the other's intention that we J (and thereby of the other's relevant actions). Each knows this while also knowing that the corresponding intention of the other participant also appropriately leads to the joint J-ing, in part by way of its support (as in (b)) for one's own relevant intention. So each can coherently see himself as having causal responsibility (but not sole causal responsibility) for the joint activity, though this causal responsibility involves other-agent conditional contribution. And this is the conclusion about causal responsibility anticipated earlier. And each can see his own intention that we J as settling whether we J while also recognizing that the other's intention also settles that. So in shared intention each can coherently intend that we J – where this intention satisfies the settle condition -- even though this intention violates the own-action condition;[140] or, at least, we have so far not been provided with a persuasive reason against this claim.

2. Interdependence in persistence and temporal asymmetry

Condition (b) involves inter-dependencies between intentions in favor of the joint activity. And these inter-dependencies are a matter of broadly rational functioning on the part of each. My recognition of your intention helps support the persistence of my intention in a way that is a matter of the rational functioning of my own psychic economy – this is not to be a case of a causal connection that is "deviant" in a way that by-passes

[140] As David Velleman points out, this conclusion is also supported by the idea – one he attributes to John Searle – that "what is intended … is whatever the state represents itself as causing and thereby tends to cause". See his review of my Faces of Intention in The Philosophical Quarterly (2001), at 121; and for the attribution to Searle see Velleman's "How to Share an Intention," at 207. This suggests that Searle's basic approach to intentionality can be seen as compatible with my view about intending that we J as ordinary intention with a special content, even though Searle himself thinks we need here to appeal to a fundamentally different attitude of we-intention. In this respect I agree with Velleman's astute observation that "Searle's account of shared intention is not entirely faithful to his own conception of what an intention is." ("How to Share an Intention," at 202).

individualistic rational functioning;[141] and similarly for you and your recognition of my intention.[142]

The idea is not that the support that each intention that we J provides for the other must be a matter of leading the other to a new intention that we J. It is possible that the onset of one intention triggers the onset of the other; though such dependence would be asymmetrical. What is needed, however, is not onset interdependence, which would be a puzzling idea, but rather that the persistence of one's intention that we J supports the continued persistence of the other's intention that we J, and vice versa. Think of two boards independently placed in a vertical position immediately next to, but not touching each other. Each continues standing only if the other does; and this is a kind of causal interdependence, since movement of one board would cause movement in the other. But for each board the causal etiology of the fact that it is standing does not involve the other board. The dependencies in (b) can be like this.[143]

For example, at the end of a wonderful concert I might intend that we applaud together. That this was a wonderful concert is, among us, what Robert Stalnaker calls a "manifest event": it is an event "that, when it occurs, is mutually recognized to have occurred."[144] This supports my confidence, given my prior knowledge about the kind of people who attend concerts like this, that you also intend that we applaud together; and so you do. Each of us arrives at his intention that we applaud primarily in response to the performance and its "manifest" quality, not to one another. Neither assures the other in the run-up to the shared applause. Nevertheless, if I were to give up my intention you would recognize this and, let's suppose, give up your intention; and this would be a matter of your own rational functioning. And vice versa. So the persistence of each of our intentions that we applaud is causally, epistemically, and rationally

[141] In unpublished work, Edward Hinchman raises a concern about deviant causal chains, and this is my response. See Edward S. Hinchman, "Regret and Responsible Agency," unpublished, dated 8/18/07. [update]

[142] Since our theoretical concern here is specifically with shared intention, we are in a position to appeal in this way, without circularity, to the rational functioning of each of the individual participants.

[143] I am responding here to issues raised by Ned Block and Christopher Peacocke in discussion and in correspondence.

[144] Robert Stalnaker, "Common Ground," Linguistics and Philosophy 25 (2002): 701-721, p. 708. I am here assuming that the 'wonderfulness' - and not just, say, the loudness -- of the concert can be "manifest".

dependent on the persistence of the other's corresponding intention, even though neither intention was part of the causal etiology of the other.

The dependence at issue here is, roughly speaking, a form of counterfactual dependence: I intend that we J; you intend that we J; if you were to cease so intending then so would I; and if I were to cease so would you. We need however to be careful about possible forms of over-determination.

Suppose that if you were to cease intending that we J you would intend something else, and suppose that this would be enough for you to make your contribution to our J-ing. And suppose that this would suffice for me to retain my intention that we J. (Though in this counterfactual situation we would no longer have a shared intention to J since you would no longer intend that we J.) In such a case it might not be true that if you were to cease intending that we J then I would also cease intending that we J. Nevertheless, in such a case what you in fact intend is that we J, and this is the intention that I recognize as in fact supporting our J-ing; and, further, I would indeed give up my intention that we J if you gave up yours and it was not replaced by a different intention that would suffice for your part in our J-ing. In this sense my intention that we J is counterfactually dependent on your intention that we J other things equal, where those other things include the absence of such a substitute intention on your part. And we will want to interpret the dependence in (b) as such a counterfactual dependence other things equal. For similar reasons we will also want this other-things-equal qualification to include other actual features of each agent's psychology that would suffice, in the absence of the intention in favor of J, to ensure that the agent plays her role in J.

In short, we will want to interpret the dependence in (b) in a way that can allow for these and other relevant cases of over-determination.

Return now to Miller's example of the two tunnel builders.[145] How might this example interact with our appeal to interdependence? Perhaps what is true in Miller's example is that I intend that we build the tunnel, and so do you; but if you were to replace your intention that we build the tunnel with an intention simply to do your part

[145] See Chapter 2 section 00.

(digging half way) I would still intend that we build the tunnel.[146] In this counterfactual situation we do not share an intention to build the tunnel together, since in this counterfactual situation (in contrast with the actual situation) you do not intend that we build it. But the question now is whether this counterfactual indicates that my intention that we build the tunnel is not appropriately dependent on your actual intention that we build the tunnel. And the answer is: No. In such a case my intention that we build the tunnel remains counterfactually dependent on your actual intention that we build the tunnel, other things equal.

There may, however, seem to be a remaining puzzle about interdependence.[147] Granted that conditions (a)-(c), when known, make it coherent that each intends that we J without either of us unilaterally determining what happens. But (b) concerns the persistence of each person's intention that we J. And we can ask: how exactly do we each arrive at those intentions? Given the interdependence in (b), it may seem that I cannot form my intention that we J until you do, but also that you cannot form your intention that we J until I do. So how could we ever arrive at these intentions?[148]

Well, though you are, we can suppose, a free agent, I can still many times reliably predict how you would freely respond to various circumstances. In particular, given my knowledge of the kind of guy you are – what you care about, how you feel about me, how you see our present circumstances, and so on -- I can many times reliably predict that if I were to intend that we J then you would know this (perhaps because I would begin acting in a way that would be best explained by such an intention), and your knowledge of my intention would lead you also to intend that we J. I can reliably predict this without supposing I have any special authority to tell you what to do, any more than I need to suppose I have such authority when I reliably predict you will tell me the time when I ask for it, or help me when I drop my package. So in many

[146] I am assuming that I can still intend that we build the tunnel even if you only intend to do your part of our building it. This is a plausible assumption, since I may know that if I so intend then you will intend your part and so together we will indeed build the tunnel.
[147] My response to this remaining puzzle is the promised second stage of my defense of condition (i).
[148] A recent version of this concern about my view is in Abraham Sesshu Roth's comments on my "Modest Sociality and the Distinctiveness of Intention," presented at the 2008 Pacific Division Meetings of the American Philosophical Association. A related discussion is in Abraham Sesshu Roth, "Shared Agency and Contralateral Commitments," Philosophical Review 113 (2004): 359-410, at 373-80. My response to follow largely derives from my discussion in "I Intend that We J" at 154-160.

cases I can form the intention that we J fully confident that you will thereby be led also to intend that we J. Given my confidence about you, I can coherently form this intention that we J. And when I do form that intention and things happen as I have predicted, we can thereby arrive at knowledge of the structure described in (a)-(c), including the relevant forms of persistence interdependence. There will then be symmetry in the persistence dependence of each person's intentions on the other, even though the etiology of those intentions involves an asymmetry.

Granted, when there is such an asymmetry in etiology, there will a period during which I intend that we J but you do not (yet). But even during this preliminary period I will believe that you will (shortly) come to intend that we J. And that is enough for me sensibly to believe, even during this time lag, that my intention that we J will be effective.

To review: the claim is that given knowledge on the part of the participants of conditions along the lines of (a)-(c), we can coherently refer, within our construction of our shared intention to J, to each person's intending that we J. We can do this even though the intentions of each violate the own-action condition (but not the settle condition). At least in the central cases we have been describing, when we share an intention to J, my intention that we J involves a known interdependence in persistence with yours. The interdependence that is central is a causal, epistemic and rational interdependence in persistence between the intentions in favor of the joint activity. This interdependence is symmetrical, though the route to these intentions may involve a temporal asymmetry. And in explaining the relevant interdependence we have made no explicit appeal to any special authority that one has to tell the other what to do, or to mutual obligations of each to the other. [149]

3. Further building blocks

In defending the coherence of both your and my intention that we J, as they appear in building block (i), I have appealed to cases in which we each know that

(a) we each intend that we J,

[149] The first feature contrasts with the view of Abraham Sesshu Roth in his "Shared Agency and Contralateral Commitments," Philosophical Review; the second contrasts with views of Margaret Gilbert.

(b) for each of us the persistence of his own intention that we J depends causally and rationally on his own continued knowledge that the other also so intends, and

(c) if we do both so intend, then as a result (and in accordance with the connection condition) we will J.

In the light of this appeal, what further building blocks should we introduce into our construction of shared intention and modest sociality?

A preliminary point is that we already have on board condition (vii), the common knowledge condition. So we can suppose that the intentions cited in (i)-(iii) are out in the open, common knowledge; and whatever further building blocks we introduce will also fall within the scope of this common knowledge.

Return to our shared intention to go to NYC. Here, in light of the appeal to (c), it seems that we will at least want to add to our construction relevant beliefs on the part of each about effectiveness:

(iv) we each believe the following: if each of us continues to intend that we go to NYC then, as a result (in a way that satisfies the connection condition), we will go to NYC.

In this way we build into the construction a condition that helps ensure the coherence of the intentions cited in conditions (i)-(iii). Granted, in order to make the case as strong as possible in defending the coherence of each participant's intention that we J, I earlier appealed to knowledge of condition (c). But now we can observe that if (assuming other relevant conditions in the background) those intentions that we J are coherent in a case in which each knows (c), they will be coherent if each believes (c).[150] So for the purpose of ensuring the coherence of relevant intentions in the construction we can just add (iv) to the construction.

A complication is that, as I have noted, I do think that it is too strong to say, quite generally, that in order to intend X one must believe X. But our primary concern now is with sufficient conditions for shared intention and modest sociality. So we can

[150] In a moment I will argue against requiring that the beliefs in (iv) be true; so it would be a mistake to render (iv) as a knowledge condition.

reasonably include among these conditions a somewhat strong belief about effectiveness – as described in (iv) -- on the part of each of the participants.

Should we also include the condition that these beliefs in (iv) are true? No. We do not want to build into our construction of shared intention a condition that says that the shared intention will actually be effective. After all, we want, in general, to allow for intentions – individual or shared – that fail. So, concerning this building block, we do best to stick with (iv), in which it is only beliefs in effectiveness that are cited – beliefs that might turn out to be false.

Turn now to (b). Here again it seems that in light of our appeal to knowledge of (b) to support the coherence of each intending the joint action, we will at least want to add to our construction relevant beliefs about interdependence:

> (v) we each believe the following: the persistence of his own intention that we go to NYC depends causally and rationally on his own continued knowledge[151] that the other also so intends; and vice versa.

Should we also include the condition that these beliefs in (v) are true? Well, I do not think we need to assume that they are true in order to guarantee the coherence of the intentions on the part of each that we go to NYC. If, as I have argued, those intentions are coherent (assuming other relevant conditions in the background) in a case in which each knows (b), they will be coherent if each believes (b). Nevertheless, it does seem that at least in central cases of shared intention and modest sociality the participants will not be in error in having these beliefs about interdependence: modest sociality will normally not be built on such an illusion. At least in central cases there will actually be something like the persistence-interdependence that is cited in the content of the beliefs in (v).[152]

As we will see below, there are important distinctions to be made here between kinds of interdependence in persistence. But the point now is only that some form of such actual interdependence in persistence is plausibly seen as a characteristic feature of shared intention and modest sociality. In the absence of some such actual

[151] In formulating this condition in this way I am exploiting the background assumption of common knowledge of the basic intentions of each.

[152] Cf. Abraham Roth: "One difference between shared activity and a mere aggregation of individual acts is that the intentions and actions of the participants in the former are dependent on one another in a manner not necessarily true of the latter." ("Shared Agency and Contralateral Commitments," at 373.)

interdependence we would have, at most, an attenuated sociality, one in which each participant's continued commitment to the joint activity is in fact not relevantly dependent on the other's corresponding commitment, though (in order to ensure the coherence of the intentions in (i)-(iii)) each believes that it is. So – given that we are looking for sufficient conditions for a robust form of sharing -- it seems reasonable to add to our anticipated construction of shared intention and modest sociality that the participants are not in error in their beliefs about this interdependence. So let us add to our construction a condition of interpersonal intention-interdependence:

> (vi) The persistence of each person's intention that we go to NYC depends causally, epistemically, and rationally on the persistence of the other's corresponding intention.[153]

The rationale for the inclusion of (vi) involves two inter-related considerations. The first is that beliefs about such interdependence are involved in each intending the joint activity. And we should expect that in normal cases of shared intention these beliefs are not in error. The second is that it is independently plausible that one characteristic feature of shared intention and modest sociality is that there is something like this interdependence between the participants.

The proposal, then, is to add versions of (iv)-(vi) to our construction of shared intention. (iv) and (v) appeal to beliefs of each that are needed to support the coherence of the intentions, cited in (i)-(iii), in favor of the joint activity. (vi) appeals to an actual interdependence between intentions of each. This actual interdependence is not itself a necessary condition of the very coherence of each intending that we J; but it is, plausibly, a central element in basic cases of shared intention and modest sociality.[154]

---

[153] Here again, as with (v), this precise formulation exploits the background condition of common knowledge, by taking it for granted that if the other's relevant intention persists one will know that, and vice versa.

[154] In moving from an appeal to knowledge of these dependencies to belief, or true belief, in these dependencies I have also put to one side whatever further condition is needed to turn true belief into knowledge. (A traditional view here is that what is needed is justification.) This seems to me right as an approach to the building blocks of shared intention. It might nevertheless in some cases be important to go on to say whether the relevant beliefs of the participants are justified; and it might be that cases in which the relevant beliefs are not justified will be non-standard in certain ways.

Now, we already have the idea that in modest sociality the intentions of each in favor of the joint activity interlock, and in that sense are semantically interconnected. This is an aspect of the building block described in (ii). And this semantic interconnection is a kind of semantic interdependence, since the success of the reference in the content of the intention of each to the intention of the other requires that the other does in fact so intend. But now we see, in condition (vi), that there will normally also be forms of causal, epistemic, and rational interdependence in persistence of the intentions in favor of the joint activity.

We also need to keep track of the difference between such semantic interlocking and the cited causal, epistemic, and rational interdependence. Suppose, on the one hand, that you and I each intend that we go to NYC. The persistence of these intentions might be interdependent even if these intentions do not semantically interlock. Perhaps each intends that they go to NYC, and would not so intend if the other did not so intend. Yet neither intends that the joint action go by way of the intention of the other.

On the other hand, suppose that our intentions that we go to NYC semantically interlock. So if you cease to intend that we go to NYC then, assuming all is out in the open, I will no longer intend that we go to NYC in part by way of your intention that we go to NYC; after all, I will know that you no longer intend that we go to NYC. But even if you cease to intend that we go to NYC I might continue to intend that we go to NYC. So there might not be persistence interdependence of the sort cited in (vi), persistence interdependence between our intentions that we go to NYC. So interlocking does not guarantee relevant persistence interdependence.

4. Forms of Interdependence

We now need to make some distinctions concerning the persistence interdependence cited in (vi), distinctions concerning both the nature of the dependence and exactly what is dependent on what.

Begin with the nature of the interdependence. The kind of interdependence that most naturally comes to mind here is one in which each person's intention in favor of the joint activity is part of the other person's reason for favoring that joint activity. Romeo, for example, might well see Juliet's intention that they flee together as part of his reason for favoring their fleeing together. Romeo's intention that they flee is

dependent on Juliet's intention that they flee (by way of Romeo's knowledge of her intention[155]) at least in part because Juliet's intention is a part of Romeo's reason for favoring their fleeing. Let us call this reason-for dependency.

Some relevant dependencies are not like this, however. A certain kind of Mafioso – call him Alex -- might favor our going to NYC but his reason for favoring this might not include that I also intend that we go: he would favor our going to NYC whatever I intended. Nevertheless, he might see that in fact we won't go unless I, as well as he, intend our going (he is not strong enough to throw me into the trunk). He sees my intention that we go as an enabling condition of our going, but not as a reason for his favoring our going. It is because he sees my intention that we go as an enabling condition of our going that he believes that it is in part because I so intend that the addition of his intention that we go will be enough to ensure that we go. And, roughly speaking, he needs the belief that this is enough to ensure that we go in order to intend that we go and not merely have a favorable attitude toward our going. He would abandon his intention in favor of our going if he were to learn that I have dropped out, since he would then believe that, in the circumstances, our joint activity is now no longer feasible. The dependency here between his intention and my intention (by way of his knowledge of my intention) is, I will say, merely an enabling dependency, not a reason-for dependency.

We can express this idea in more general terms. Suppose that S intends X. This intention has Y as an enabling condition when Y is needed, in the circumstances, for S's intention effectively to issue in X. And Y may be such an enabling condition without its being part of S's reason for favoring X. So we will want to distinguish enabling from reason-for dependencies.[156]

That said, it seems to me that either sort of dependency between intentions that we J is enough for condition (vi). Alex does intend that we go to NYC together, and -- if I too am in the right frame of mind, and if all the other relevant conditions are satisfied -- might be a partner in our shared intention to go. Granted, in the kind of case

[155] Here again condition (vii) is in the background.

[156] This distinction is close to the distinction between enabling conditions and reasons I once drew in discussing Donald Davidson's account of conditional intention. See my "Davidson's Theory of Intention," in my Faces of Intention at 217-19.

suggested by the example, Alex may not satisfy the further conditions for shared intention, including, for example, intending that my intention be effective. And granted too that in many cases of shared intention these dependencies will be, in contrast, reason-for dependencies. That I intend that we go to NYC will, many times, be part of your reason for favoring our going to NYC: in this respect you are more like Romeo than Alex. But I will not insist on this as a condition on each shared intention.

It is true that if we each intend that we J, if the other's intention functions for each only as an enabling condition, and if we satisfy the other conditions for shared intention, then our sharing will involve a kind of opportunistic attitude of each toward the other. Each is prepared, as it were, to take advantage of the other's intention in favor of their J-ing, but does not treat the other's intention as a reason for favoring their J-ing. This is different from the normal attitudes towards each other of lovers or friends. Still, not all sharing is the sharing of lovers or of friends; and such opportunistic attitudes do not seem to baffle shared intention so long as the other conditions for shared intention are satisfied. Modest sociality can be opportunistic sociality.

Such modest but opportunistic sociality is not merely a matter of strategic interaction, since even though the sociality is opportunistic each intends that the activity proceed appropriately by way of the relevant intention of the other. Even if I know that the dependence of my intention on yours is only an enabling dependence I can still (in contrast with certain mafia cases) intend that our going together to NYC involve the effective role of your intention. I can be disposed to help you play your role if that help is needed. And I can be set to filter options with an eye to consistency with the role of your intention in our so acting. After all, even if I do not see your intention as part of my reason for favoring our joint activity, I do see that your intention and its effectiveness is, in the circumstances, needed for us so to act. And that can be a reason for me to intend that your intention be effective, as an element in our joint activity, and not merely to expect that you will act in accordance with what you in fact intend.

Having distinguished between reason-for and enabling dependencies, we now need to make some distinctions concerning precisely what is dependent on what. Let us suppose that you and I each intend that we J, and also that you and I each intend the specific actions that would be his part of J – where this need not involve intending those

actions as part of our J-ing. (For simplicity I will assume there is a unique set of such specific acts on the part of each, and that we each already intend them.) So we have:

1A. I intend that we J

1B. I intend the acts that would be my part of J

2A. You intend that we J

2B. You intend the acts that would be your part of J.

I assume that, for each of us, if he intends the joint activity then he intends those acts that would be his part of that joint activity. However, we cannot assume the converse in either case, since one might intend those acts that would be one's part in J without intending J. (One might have a different rationale for those acts.) Bracketing the issue of whether the dependencies are reason-for or merely enabling, we can then ask which of my intentions (1A and/or 1B) are dependent on which of your intentions (2A and/or 2B), and vice versa. To keep things simpler let's assume symmetry: in whatever way my intentions depend on yours, yours depend on mine. With this assumption of symmetry in hand, we can focus on the dependency of my intentions on yours. And here there seem to be four different (mutually compatible) ways in which my intentions may depend on yours, as indicated in the following matrix:

	2A. You intend that we J	2B. You intend the acts that would be your part of J.
1A. I intend that we J	I.	II.
1B. I intend the acts that would be my part of J	III.	IV.

In I-dependence my intention that we J depends on your intention that we J. In II-dependence my intention that we J depends on your intention to perform the acts that would be your part of J. In III-dependence my intention to perform those acts that would be my part of J depends on your intention that we J. And in IV-dependence my intention to perform those acts that would be my part of J depends on your intention to perform the acts that would be your part of J.[157]

[157] IV-dependence is close to the kind of dependence to which David Lewis confines his attention. See his definition of a coordination problem in his Convention at p. 24.

Condition (vi) is a condition of I-dependence: each person's intention that we J depends on the other's intention that we J. And the conjecture is that this kind of dependence is central to modest sociality. But in many—though not all -- cases of modest sociality there will also be some or all of the other modes of dependence.

Suppose that we share an intention to push the piano up the stairway. I intend that we push it, you intend that we push it, and these intentions in favor of the joint pushing are I-interdependent. We can also expect that I would not intend that we move the piano if you did not intend to do your part; and vice versa. So there is II-interdependence. Further, we may suppose that it is also true that each of our intentions to do what would be his own part is dependent on the other's corresponding intention to do what would be his own part. I would not intend to push if I did not believe that you intended to push. This would be IV-interdependence. And we might also suppose that I would not continue to intend to do my part if I discovered that you had given up on your intention in favor of our joint piano moving. Perhaps I find such strenuous effort on my part too unattractive unless it is a part of a joint activity that we each intend. This would then be III-dependence.

So in a case of modest sociality like this we have interdependence of all four sorts. Further, it seems that each of these forms of interdependence might either be reason-for or enabling.[158] Let's say that when the interdependence is of all four sorts – I through IV -- the interdependence is extensive. And now the point to note that in some cases of modest sociality the interdependence will not be extensive.

Return to our shared applause.[159] My intention in favor of our applause depends on your intention in favor of our applause. This is I-dependence. Further, I would no longer intend our applause if I knew that you did not even intend to applaud. So this is

[158] A dependence of my intentions to perform what would be my part – III- or IV-dependence -- will frequently be reason-for dependence. That you too intend to push the piano gives me grounds for thinking that my pushing the piano would be a safe and effective thing to do, and so helps provide what I see as a reason for my pushing. But a dependence on your intentions of my intention to perform what would be my own part can sometimes be only an enabling dependence. This can happen if my very ability to perform what would be my part depends on your performing your part. Perhaps I could not even move my body in a tango-appropriate way if you did not move your body in a tango-appropriate way – just as I cannot move my body in the way involved in tying my tie, or my shoelaces, without the presence of my tie, or my shoelaces. So my ability to perform my part of our tango depends on your performing your part. So this would be a case of IV (and perhaps III) dependence that is an enabling dependence.

[159] See above, section 2.

II-dependence. But now suppose that were I to recognize that you have given up your intention in favor of our applause I would nevertheless still intend to applaud on my own: this is not III-dependence. Relatedly, suppose that even if I were to recognize that you had no intention to do what would have been your part in our shared applause (that is, applaud) I would still intend to do what would have been my part: I would still intend to applaud. This is not IV-dependence.

Again, suppose that both Romeo and Juliet each have reasons to flee that do not depend on a reason that involves the other's intention in favor of their fleeing. Romeo, let us say, is being chased by the tax collector and so has a reason to flee that does not depend on Juliet's intention that they flee, though he also has a reason to favor their joint fleeing that does depend on that intention of Juliet's. So Romeo is prepared to flee whether or not Juliet does. So while Romeo's intention in favor of the joint fleeing depends on Juliet's intention favor of their fleeing –that intention of Juliet's is a crucial reason of his in favor of a joint fleeing -- Romeo's intention to perform what would be his own part (if they were to flee together) does not depend either on Juliet's intention that they flee or on Juliet's intention to do what would be her part in their fleeing. So this is I-dependence without III- or IV-dependence.

So there can be cases of I and II dependence without III or IV dependence. Let's say that such dependency is limited dependency. As indicated by these examples, our construction of modest sociality will want a version of condition (vi) that allows both for extensive interdependence and for limited interdependence. And in both cases it will want to allow both for reason-for and for enabling dependencies.

5. The connection condition and mutual responsiveness

In shared intentional activity joint action is appropriately explained by a relevant shared intention. I have called the condition that specifies the nature of this explanatory relation the connection condition. Appeal to this connection condition is built into the contents of attitudes cited in the condition of interlocking and reflexive intentions in (ii), and the beliefs about efficacy in (iv). And this connection condition is itself part of the metaphysics of shared intentional activity (though, as noted, we want to keep open the possibility of unsuccessful shared intention). We need then to say more about what this connection condition involves.

Suppose there is a shared intention that issues in corresponding joint action. What can we say about the connection between that shared intention and the joint action when this joint action is thereby a shared intentional action?

My conjecture is that the standard route from our shared intention to our joint action, in a case of shared intentionality, involves relevant mutual responsiveness of each to each in a way that tracks the intended joint action, where this mutual responsiveness is out in the open. There will be responsiveness of each to each in relevant subsidiary intentions to perform means and preliminary steps. This is responsiveness in intentions to act that are elements in sub-plans concerning the intended joint action; and a tendency toward this is normally supported by the condition that the intentions in favor of the joint activity interlock and that each intends that the joint activity proceed by way of sub-plans that mesh. There will, further, be responsiveness of each to each in relevant actions in pursuit of the joint activity; and a tendency toward this is also normally supported by the condition that the intentions in favor of the joint activity interlock and favor mesh. This is responsiveness in action. Responsiveness in action is largely, though perhaps not entirely shaped by responsiveness in sub-intentions. However, in appealing not only to responsiveness in sub-intentions but also to responsiveness in action we are making it explicit that the relevant responsiveness goes all the way to actual action.[160]

The mutual responsiveness that is relevant here is in the space between two extremes. On one extreme there is on the part of each a very general responsiveness to and support of the aims of each of the others. As we might say, in a broadly Kantian spirit, each treats the other as end-providing. In contrast, the mutual responsiveness that is involved in the connection condition is specifically limited to the particular joint activity that is intended (though of course it does not preclude a more general responsiveness). On the other extreme is a mutual responsiveness in which each is pursuing an end incompatible with the end that is being pursued by the other. This is the mutual responsiveness of two soldiers who are enemies and are fighting each other outside of a context of a shared intention. In contrast, the mutual responsiveness that is involved in the connection condition tracks the joint action that each intends. So the

[160] Luis Cheng-Guajardo highlights closely related ideas in unpublished work.

mutual responsiveness that is involved in the connection condition is both relativized specifically to the intended joint activity and tracks that intended joint activity. In this way it is in the space between an extremely general responsiveness and a responsiveness that does not track a common end.

In appealing in this way to mutual responsiveness I am presupposing that the connections between your intentions to act and your actions are appropriate for the individual intentionality of your actions; and similarly concerning the individual intentionality of my actions. My question here is: in shared intentional action what is normally present, in the connection between our intentions and our joint action, that goes beyond these basic connections between each person's thought and that person's individual intentional actions? And my answer is: responsiveness of each to each both in sub-intention and in action, responsiveness that is relativized to and tracks the intended joint action, and responsiveness that is out in the open.

Return to our duet singing. The idea is that if our singing the duet is a shared intentional activity that is grounded in our corresponding shared intention, then my sub-intention about, say, when to come in with my own part, and my associated actions, will be responsive to your sub-intention about your own part, and your associated actions, in ways that track our joint duet singing. And vice versa. There can be such mutual responsiveness even if in fact no actual adjustment takes place since none is called for. However, if there is such mutual responsiveness then each is at least set to adjust sub-plans and actions appropriately, and the situation is one in which each has some sort of cognitive access to conditions that would call for adjustment.

My claim, then, is that for shared intention to lead in the right way to shared action is for the process to involve such mutual responsiveness of sub-intention and in action, mutual responsiveness that is relative to and tracks the intended joint activity.

This may seem overstated. After all, there will not be such mutual responsiveness in action if we decide in advance on our respective roles and then simply proceed to act individually in a prepackaged way that does not involve even the possibility of mutual adjustment of each to each. Two explorers might set out in different directions with a prior shared plan about what each will do, knowing that there is no possibility of contact once they begin (there are no cell phones), and so no possibility of

mutual adjustment and responsiveness once they begin. So we need to ask whether a condition of mutual responsiveness is overly demanding.

Well, such cases of prepackaging are an analogue, in the shared case, of a kind of ballistic action in the individual case. Individual intentional action normally involves a kind of responsiveness of action to the world, responsiveness that tracks what is intended. If I intend to guide the boat into the harbor I will normally adjust in response to the various currents and in a way that tracks the harbor. Such responsiveness seems a normal feature of the connection between intention and action, a feature that is characteristic of individual intentional action. But there can be cases of intentional action in which one simply exerts an immediate, one-off effort, and there is no further room for responsiveness: one acts and then the rest is up to the world. Perhaps I get to push the boat just this once, and the rest is up to the currents – just as I throw a bowling ball down the alley with an eye on the top pin. Prepackaged shared action is a shared analogue of such ballistic action. In both cases we have intentional action – individual or shared – but we do not have the cited forms of downstream responsiveness.

That said, we do have a more robust phenomenon of shared intentionality when there is the cited kind of mutual responsiveness in sub-intention and action. And our basic concern is with the question whether we can get robust shared intentionality without retreating, in the end, either to a doctrine of the conceptual primitiveness of shared intentionality or to a doctrine of a deep metaphysical or normative discontinuity between individual and shared agency. So there is reason to build the stronger condition of mutual responsiveness in sub-intention and action into our construction while acknowledging the possibility of attenuated forms of shared intentionality in the absence of such mutual responsiveness.

I will proceed, then, on the assumption that the connection condition can be analyzed in terms of such mutual responsiveness in sub-intention and action, even though I grant that there can be attenuated forms of shared agency of the prepackaged variety. Insofar as intentions and beliefs of the participants appeal in their contents to this connection condition – as they do in conditions (ii) and (iv) -- we can understand this as an appeal to such mutual responsiveness. Since we have analyzed this condition of mutual responsiveness without an irreducible appeal to the very idea of

shared intentionality, we can include this appeal to mutual responsiveness in these contents without an unacceptable circularity in our account of shared intentionality. And that is what we wanted to be in a position to do.[161]

This mutual responsiveness is a form of interconnection in functioning between the intentions and actions of the participants. This interconnected functioning across the intentions and actions of each is central to the standard connection between social thought and social action. This interconnected functioning is, as noted earlier, normally explained in large part by the relevant intentions of those participants. It is in large part because of the inter-locking intentions of each in favor of the joint action by way of meshing sub-plans that each is set to be relevantly responsive to the other in downstream thought and action.

In being normally explained by these inter-locking intentions, such mutual responsiveness in sub-intention and action contrasts with the interdependence in persistence of intention cited in (vi). There is no claim that this interdependence in persistence of intentions in favor of the joint activity is normally explained by appeal to the intentions of each. Indeed, in a way the dependency goes in the opposite direction: it is in part because the agents each believe there is this interdependence in persistence of intention in favor of the joint activity that they can coherently intend the joint activity. That said, it remains possible that each does go on to intend that there be this interdependence in persistence; and such a further intention might be an important aspect of certain cases of modest sociality.

This may seem to be a tension with our understanding of condition (ii). That condition – as we are now understanding it – says that each intends that the joint activity proceed by way of the corresponding intention of the other and relevant mutual responsiveness of sub-intention and in action. Does this entail that each intends that the intentions of each in favor of the joint activity are interdependent in the sense of condition (vi)? No, it does not. The intended mutual responsiveness cited in (ii) (as we are now understanding the appeal in that condition to the connection condition) concerns each participant's relevant sub-intentions: it is responsiveness in each participant's sub-plans with respect to the intended joint activity. One can (as in (ii))

[161] See Chap. 2, section 2.

intend that things go by way of such responsiveness in sub-plan without intending that (as in (vi)) the intentions in favor of the end of the joint activity themselves be interdependent.

We can now return to a question that was raised earlier about the relation between (ii) and (iii).[162] Suppose that one has the intention cited in (ii). Does it follow that one intends that the joint activity proceed by way of meshing sub-plans? Now that we have an account of the connection condition that is cited in (ii), an account that appeals to relevant mutual responsiveness, we can answer this question.

Suppose, as in (ii), that I intend that we act by way of your intention that we act and my intention that we act, and by way of a route from our intentions to our joint activity that satisfies the connection condition. As we are now understanding it, this is an intention that we act by way of relevant mutual responsiveness in sub-intention and action. This intended mutual responsiveness includes (but is not limited to) mutual responsiveness in the sub-intentions that constitute the relevant sub-plans of each. So the intention in (ii) involves an intention that these relevant sub-plans be mutually responsive to each other. And that amounts to an intention that these sub-plans mesh. The intention that the activity proceeds by way of sub-plans that mesh is, then, one aspect of the intention that that activity proceed by way of mutual responsiveness: intending that there be mutual responsiveness in – in particular -- sub-intention involves intending that there be mesh in sub-plans. (Though, as noted earlier, the intention that there be responsiveness in action is not just an intention that there be responsiveness in sub-plan.)

So when we have a full understanding of condition (ii) – one that draws on our substantive account of the connection condition – we can see that (iii) is indeed implicit in (ii). So we can see (ii), suitably understood, as in part the condition that each intends the joint activity by way of the other's intention, mutual responsiveness, and so meshing sub-plans. And this helps clarify the significance of the intention in favor of mesh in sub-plans. However, though (iii) is implicit in (ii) there remain good reasons for explicitly highlighting (iii) as a condition of modest sociality, given its prominent role in our

[162] Chap. 2, section 3.

construction. (This is parallel to what I said earlier in discussing a similar redundancy between (ii) and (i).) So that is what I will normally do.

This completes my discussion of the basic building blocks for our construction. Now let's put them together.

Chapter Four: A Construction of Shared Intention and Modest Sociality

1. The basic thesis

We now have at hand the basic building blocks we need for a conservative construction of modest sociality. Keeping in mind the focus on sufficient conditions, the first idea will be to see generalized and adjusted versions of (i)-(vii) as together constituting shared intention (or, anyway, an important form of shared intention). And the second idea will be to see modest sociality (or, anyway, an important form of modest sociality) as joint activity that is appropriately explained – in part by way of mutual responsiveness -- by such shared intention.

Consider first shared intention. Using **boldface** to indicate adjusted generalizations of the conditions cited earlier, the idea is that shared intention, at least in a basic case, involves

- **(i)** non-attenuated intentions on the part of each in favor of the joint activity,
- **(ii)** interlocking and reflexive non-attenuated intentions, on the part of each, in favor of the joint activity by way of relevant mutual responsiveness in sub-intention and action,
- **(iii)** non-attenuated intentions on the part of each in favor of the joint activity by way of meshing sub-plans of the intentions in **(i)**,
- **(iv)** beliefs of each about the joint efficacy, by way of mutual responsiveness in sub-intention and action, of the intentions in **(i)**,
- **(v)** beliefs of each about interpersonal causal, rational, and epistemic intention-interdependence in persistence of the intentions in **(i)**,[163]
- **(vi)** interpersonal causal, rational, and epistemic interdependence in persistence of the intentions in **(i)**, and
- **(vii)** common knowledge of **(i)-(vii)**.

Since the relevant intentions are non-attenuated, these intentions will involve associated **dispositions to help if needed**. These intentions of the individual participants engage

[163] As noted, the persistence interdependence in **(vi)** can take various forms. Since our interest is in sufficient conditions, I will assume that the beliefs about the interdependence, cited in **(v)**, accurately track the form of interdependence that is present, though we could weaken this condition if that seemed appropriate.

characteristic **norms** of individual intention rationality: norms of consistency, agglomerativity, coherence, and stability. And these norms have a distinctive normative significance.

Consider now the connection between shared intention and joint action that is characteristic of shared intentional activity. This **connection condition** is satisfied just in case:

> **(viii)** the shared intention (as in **(i)-(vii)**) leads to the joint action by way of relevant mutual responsiveness in sub-intention and action that tracks the intended joint activity, in conditions of common knowledge.

(viii) is a condition on the actual interactions in the functioning of relevant intentions. However, these interactions will themselves normally be explainable by appeal to the intentions of those participants. It is normally because each intends that the joint activity go by way of the intentions of each, mutual responsiveness, and so meshing sub-plans, that each is set to be responsive to the other in relevant ways. (Of course, there may be failure.)

In articulating these conditions I have not put a great deal of weight on appeals to our ordinary talk about shared intention, shared intentionality, and the like. Indeed, one of my reasons for introducing the term "modest sociality" has been to help us focus on the actual phenomena, not primarily on ordinary language. I do think that the model I am sketching broadly coheres with pre-analytic talk of what we intend, and of our shared intentional and shared cooperative activities. And such pre-analytic talk can be a useful, if defeasible, guide. But my primary concern is not with our pre-analytic talk but with shared intention as a central element in the explanation of the activities involved in (what I am calling) modest sociality. I take it that in the case of individual intentional agency, intentions are central to the explanation of the agent's relevant practical thought and action. And I seek a model of shared intention that helps us get at a phenomenon that plays an analogous explanatory role in the case of modest sociality. In each case I suppose that a fundamental kind of human activity – temporally extended activity, on the one hand, and, on the other hand, shared intentional activity -- essentially involves a distinctive explanatory role of relevant aspects of mind. And in each case my conjecture is that this explanatory role involves capacities of planning agency.

Returning to conditions **(i)-(vii)**, the central claim is that these inter-connected intentions and associated beliefs of the individual participants, all in a context of common knowledge, will, in responding to the rational pressures specified by the planning theory of individual agency, function together in ways characteristic of shared intention. This structure will, when functioning properly, normally support and guide coordinated social action and planning, and frame relevant bargaining and shared deliberation, in support of the intended shared activity. Conformity to social rationality norms that are central to shared intention -- norms of social agglomeration, social consistency, social coherence, and social stability -- will emerge from the norm-guided functioning of these interrelated attitudes of the individuals. Violation of such social norms will normally consist in a violation of associated norms of individual planning agency, a violation that will constitute a rational breakdown on the part of that individual. And when such structures of shared intention explain our activity by way of the connection condition -- condition **(viii)** – our activity is a shared intentional activity and a candidate for shared cooperative activity.

Call the conjunction of these claims about shared intention and modest sociality the basic thesis. The thesis is that shared intention and modest sociality consist, at least in central cases, in appropriately inter-related structures of individual planning agency, in a context of common knowledge. These inter-related planning structures go beyond the merely cognitive inter-relations involved in knowledge of each other's minds and present in standard forms of merely strategic interaction. But these structures go beyond these merely cognitive inter-relations in ways whose understanding involves the application of conceptual, metaphysical and normative resources that are available within the theory of individual planning agency. Such shared intention consists in a complex state of affairs that involves relevant, interconnected attitudes of individual planning agents. And modest sociality consists in the proper functioning of such shared intentions. Such modest sociality is inter-connected planning agency.

The basic thesis provides a model of the social glue that ties together the participants in modest sociality.[164] According to this model, this social glue is not solely a cognitive glue of belief and common knowledge, though it does involve a form of common knowledge. This social glue also includes the forms of intentional interconnection specified in (**ii**)-(**iii**), dispositions of inter-personal support implicit in **(i)-(iii)**, beliefs about joint efficacy and interdependence cited in **(iv)** and **(v)**, the actual interdependence in persistence cited in (**vi**), and the mutual responsiveness in sub-intention and action cited in **(viii)**. The forms of intentional interconnection cited in **(ii)-(iii)** involve semantic interrelations across the intentions of the different participants, semantic interrelations similar to those we have observed to be characteristic across the plan states of an individual at different times in her temporally extended planning agency. These interrelated intentions in favor of the joint activity themselves help support mutual responsiveness in sub-intention and action in the pursuit of that joint activity, mutual responsiveness of the sort cited in **(viii)**. And the intention interdependence cited in **(vi)** involves actual causal, epistemic, and rational interrelations across the intentions of the different participants.

2. The emergence of modest sociality

Suppose then that there is the structure of inter-related intentions, beliefs, and common knowledge cited in **(i)-(vii)**. And suppose these attitudes of the individuals function properly in the sense spelled out by the theory of individual, rational planning agency. Given the contents of these intentions there will tend to be the mutual responsiveness cited in **(viii)**. And when all this works its way out, without interference or breakdown, there will be the social functioning and social rationality that is characteristic of modest sociality. Let us see in more detail why this is so.

Condition **(i)** helps ensure that each participant to some extent rationally tracks not only his or her actions but also the joint activity – where this will normally include associated dispositions to help the other if needed. And condition **(i)** also helps ensure that each participant rationally filters options for deliberation with an eye on compatibility with the joint activity. Conditions **(ii)** and **(iii)** help articulate in a more fine-grained way

[164] For talk of "the glue that binds team members together" see Philip R. Cohen, Hector J. Levesque and Ira A. Smith, "On Team Formation," in G. Holmstrom-Hintikka and R. Tuomela, eds., Contemporary Action Theory vol. II (Dordrecht: Kluwer Academic Publishers, 1997): 87- 114, at 89.

what each participant tracks. And they thereby help to explain characteristic ways in which each tends to support the role of the other. They thereby go beyond merely cognitive links among the participants to capture an important way in which each is treated as a co-participant.

Conditions **(iv)** and **(v)** ensure that the participants see themselves as both inter-dependent and, together, practically effective – and thereby help support the coherence of the intentions of each in favor of the joint activity. Condition **(vi)** captures an actual interdependence between the agents, an interdependence that can take various forms. The common knowledge condition – condition **(vii)** -- helps ensure that the construction of shared intention will itself support a standard form of functioning associated with modest sociality, namely: thinking (including thinking together) about what to do given that we share this intention. This common knowledge condition also helps ensure -- by ensuring that the basic intentions of each are out in the open -- that the believed inter-dependence in **(v)** concerns the actual intentions of each. And the mutual responsiveness condition – condition **(viii)** -- captures the connection between social thought and social action that is characteristic of modest sociality, a connection that is supported by the inter-connected intentions cited in **(i)-(iii)**.

Consider then my intention that we J in part by way of your analogous intention, mutual responsiveness and so meshing sub-plans. This complex content of my intention connects it with your intentions and thereby imposes rational pressure on me, as time goes by, to fill in my sub-plans in ways that, in particular, fit with and support yours as you fill in your sub-plans. This pressure derives from the rational demand on me to make my own plans means-end coherent and consistent, given the ways in which your intentions enter into the content of my intentions. By requiring that my intention both interlock with yours, and involve a commitment to responsiveness to and mesh with yours, the theory ensures that rational pressures on me to be responsive to and to coordinate with you – rational pressures characteristic of shared intention -- are built right into my own plans, given their special content and given demands of consistency and coherence directly on my own plans. And similarly with you. So there will normally be the kind of mutual, rational responsiveness in intention -- in the direction of social

agglomeration, social consistency, and social coherence -- that is characteristic of modest sociality.

Suppose, for example, that I intend that we paint together in part by way of your analogous intention, mutual responsiveness and so meshing sub-plans. Given the way reference to your intention, and to responsiveness and mesh, appears in the content of my own intention, I am committed to filling in my sub-plans in a way that responds to how you fill in yours; and I am committed to being responsive to what is needed for your analogous intention to be effective and for our intentions together to be effective. And similarly with you, assuming you too intend that we paint in part by way of my analogous intention, mutual responsiveness and meshing sub-plans. So what emerges from these intentions, given guidance by central norms of individual planning agency, are forms of mutual responsiveness characteristic of modest sociality.

The basic thesis works, in part, by building appropriate reference to the other into the contents of the intentions of each. It does not just appeal to "a Background sense of the other as a candidate for cooperative agency".[165] It seeks to generate relevant rational normativity, and corresponding functioning, at the social level out of the individualistic normativity, and corresponding functioning, that is tied to the specific (though perhaps only implicit) contents of the intentions of each. The basic thesis seeks contents of the intentions of each that ensure, given rational demands on those intentions of each (demands rooted in the planning theory of individual agency), responsiveness to central social rationality demands.

This explains how the basic thesis sees the social norms of consistency, agglomerativity and coherence, and corresponding functioning, as emerging from associated individualistic norms and corresponding functioning. This emergence of these social rationality norms depends on the presence of relevant intentions of the part of each, but does not depend on exactly what reasons keep those intentions in place.

What about the social norm of diachronic stability? Well, there will frequently be a social version of the snowball effect: once we are embarked on our shared activity there will frequently be as a result new reasons to continue. But the basic thesis also

[165] Searle (1990, 414)

helps explain a kind of diachronic rational stability of shared intention that goes beyond this social snowball effect. The intentions of each of the participants are themselves subject to characteristic rational pressures for stability of the intentions of individuals. And the different forms of interdependence in persistence characteristic of shared intention will help extend this stability of the intentions of each. If your intention that we J depends on mine and tends to persist so long as mine does, and mine is stable in part because of the constraints of a norm of stability, your intention will to some extent inherit this stability. And vice versa. So there will be a kind of mutual, rational stabilization of the intentions in favor of J, a stabilization that emerges from the rational pressures for stability of the intentions of each, together with the different forms of interdependence in persistence of those intentions.[166] This helps to explain why shared intention involves a kind of homeostatic rational stability. And this can explain a certain kind of rational criticism of certain cases of opting out of a shared intention.

Or anyway, this is the primary way in which the basic thesis explains the emergence of these social norms of intention rationality within structures of individual planning agency with appropriate contents and inter-relations. In Chapter 8, I will be discussing yet a further possible form of support for these social norms, a form of support that derives from certain kinds of shared policies in favor of these social norms (where such shared policies will themselves be understood by way of the basic thesis).

Consider now the way in which a shared intention can frame bargaining about means. Return to our shared intention to paint together. Given this shared intention we might, for example, bargain about what color to use, and about who is to scrape and who is to paint; and this bargaining will be framed by our shared intention. How does this work? Well, on the theory, we each intend the shared activity in part by way of the intentions of the other and by way of meshing sub-plans. Each of our intentions involves intending that the other's intentions be effective by way of sub-plans that mesh inter-personally. So we are each under rational pressure to seek to ensure that our sub-plans, agglomerated together, both are adequate to the shared task and do indeed

[166] Facundo Alonso independently discusses and endorses this idea – as he calls it, a "virtuous circle of mutual reinforcement" -- in important detail his Shared Intention, Reliance, and Interpersonal Obligation (Ph.D. Thesis, Stanford University, 2008), chapter 7.

mesh inter-personally, though of course we might fail. And that is why, in the absence so far of adequate and meshing sub-plans, our shared intention will tend to structure our bargaining in the pursuit of such sub-plans.[167] Of course, the mere shared intention, by itself, may not suffice to resolve the problem of settling on meshing sub-plans.[168] Various further cognitive and conative resources will frequently be needed. Nevertheless, the shared intention is a source of rational pressure to try to use such further resources to solve this problem. And we explain that rational pressure as emerging from the rational pressures on individual planning agency, given these distinctive contents of relevant intentions.

The sharing of intention need not require commonality in each agent's reasons for participating in the sharing. You and I can have a shared intention to paint the house together, even though I participate because I want to change the color whereas you participate because you want to remove the mildew. Or you and I can have a conversation together even though I participate because I want to learn more about the subject of our conversation, whereas you participate because you want to impress me.[169] In each case, though we participate for different reasons, our shared intention nevertheless establishes a shared framework of commitments. Granted, extreme divergence in background reasons might undermine the shared intention. Things might fall apart. Nevertheless, much of our sociality is partial in the sense that it involves sharing in the face of divergence of background reasons for the sharing.

Indeed, I think that just as partiality in plan is common within our individual planning agency, versions of such social partiality are endemic to our modest sociality. Such social partiality is not just partiality of social plans of action; it also includes partiality in convergence in relevant background reasons for which the different parties participate. It is a basic fact about our sociality that we manage to share intentions and act together in the face of substantial differences of reasons and of view. We work together, we play together, and we engage in conversations together even given substantial background differences in our reasons for participation and our reasons for

[167] As indicated in Chapter 1, a related but different phenomenon is shared deliberation in the pursuit of mesh. I discuss these matters in Chapter 8.
[168] As Luca Tummolini has emphasized.
[169] An example from Christopher Potts.

various sub-plans. This is especially characteristic of a pluralistic, liberal culture. Let's call this the pervasiveness of partiality in our sociality. It is a virtue of the basic thesis that it makes room, in a theoretically natural way, for this pervasiveness of partiality in our sociality.[170] I return to this idea in Chapter 8.

Shared intentions coordinate planning and action and frame relevant bargaining, all this in ways that are shaped by characteristic norms of social rationality. And what we have seen is that this complex of social functioning and social rationality will emerge from the individualistic structures described by the basic thesis. Shared intention also structures shared deliberation. However, I will postpone to Chapter 8 the project of explaining how this characteristic social functioning of shared intention also emerges from such individualistic structures. Once this promise is fulfilled we will have what we wanted: an account of the emergence of modest sociality from the individualistic structures described by the basic thesis.

3. Further reflections on the basic thesis

In its understanding of modest sociality, the basic thesis appeals not only to planning attitudes in the heads of each of the individuals but also to important inter-relations among those attitudes of the different participants. It sees modest sociality as a special form of inter-connected planning agency. This contrasts with John Searle's apparent assumption that what is essential to shared intentional agency is exhausted by certain attitudes in the heads of the involved participants.[171] However, though the basic thesis makes essential appeal to inter-relations across the participants, it aims to

[170] So I reject the thought, from Raimo Tuomela, that the agents in the kind of sociality of interest here, and characteristic of a world of multiple and diverging background reasons and values, "have their intentions of the form 'I intend that we 'J (for each participant) necessarily because of a group reason…" Raimo Tuomela, The Philosophy of Sociality: The Shared Point of View (Oxford University Press, 2007) 100-101. I grant, however, that there is pressure in the direction of convergence of background frameworks: this is a topic of Chapter 8.

[171] Searle writes: "all intentionality, whether collective or individual, could be had by a brain in a vat…".("Collective Intentions and Actions," 407). Strictly speaking, this is not yet to say that modest sociality – where this includes shared intentional and shared cooperative activity – is solely a matter of what could be had by a brain in a vat. But since Searle's entire theory of "collective intentionality" is a theory of the we-intentions that could be had by a brain in a vat, it seems that he at least implicitly endorses this stronger thought. And it is this stronger thought that the basic thesis rejects. (John Hund also reads Searle in this way, and makes a similar point in his "Searle's construction of social reality," Philosophy of the Social Sciences 28 (1998): 122-31, at 129. [Thanks to Facundo Alonso for this reference.])

characterize these interrelations without an essential appeal to either distinctive kinds of inter-personal obligation or special forms of asymmetric authority.

Return now to the point, from Chapter 1, that the normal functioning of a planning system in temporally extended individual agency systematically involves cross-temporally stable and referentially interconnected plan-states. We noted there that such stability and interconnection involves cross-temporal ties that are, on a broadly Lockean view, aspects of the persistence of one and the same person over time. What we have now seen is that the basic thesis highlights referential interlocking of the intentions of the different participants in modest sociality, and commonality of content of their relevant intentions. In this limited respect, then, the basic thesis sees a parallel between the social glue characteristic of modest sociality and the cited Lockean structure of individual planning agency over time. So we can say that, according to the basic thesis, these social ties are quasi-Lockean.

We need to understand this idea with care, however. The parallel is that in both the case of a person over time and in the case of modest sociality there are referential interconnections and characteristic commonalities. These interconnections and commonalities in the shared case help support the claim that there is an important kind of shared agency, not just a concatenation of the agency of each. But of course there are also differences. These ties in modest sociality will normally be quite limited both in their extent and in their duration: you and I might push the piano up the stairway just for a couple of minutes and then go our separate ways; and even while we are pushing we might each be independently doing various other things – rehearsing a different poem in each of our heads, perhaps. In contrast, the ties that are characteristic of the persistence of one and the same person over time will normally involve a rich overlay of ties both at each time and over time. Further, in modest sociality these ties will normally cross-cut. I might be walking with you while I am having a conversation with someone else on my cell phone; and in such a case there are, so to speak, two "we's," not one.[172]

Yet a further limitation to the parallel concerns the role of consciousness. Locke famously claimed that what is central to being a person is that a person "can consider itself as itself, the same thinking thing, in different times and places; which it does only

[172] In what sense more precisely are there two "we's"? That is the subject of Chapter 7.

by that consciousness which is inseparable from thinking, and, as it seems to me essential to it…"[173] Following David Velleman, we can express a central aspect of this remark as the idea that there are certain experiences in the future that "we can anticipate first-personally",[174] and certain experiences in the past that we "can think of reflexively, in the first-person."[175] In each case one has a special kind of first-personal experiential access to the point of view of a past or future subject of experience. One central aspect of a Lockean view of personal identity over time will be that the relevant cross-temporal ties include such first-personal experiential access to the point of view of a past or future subject of experience. And this first-personal experiential access goes beyond the causal, semantic and epistemic connections highlighted by the parallel we have noted between the intra- and inter-personal cases. I can anticipate, or remember certain experiences of mine "first-personally"; but in acting together with you I do not thereby come to be able to anticipate or remember "first-personally" relevant events in your life.

So while I think it is an important fact that the ties characteristic of modest sociality are quasi-Lockean in the indicated way, there are also important dis-analogies between these ties and the Lockean ties characteristic of personal identity. I will return to this point in Chapter 7, where I will explain why the basic thesis is not committed to there being a group subject of a shared intention.

As we have seen, it is tempting in theorizing about shared agency to suppose that there is, at bottom, some sort of conceptual, metaphysical, or normative irreducibility to structures of individual agency. And it is tempting to suppose that an effort to provide some such reductions will violate certain coherence constraints on the concepts involved – this is what is at stake, for example, in the worries about the violation of the own action condition on the content of intention. In contrast, a conservative constructivism aims to provide coherent sufficient conditions for shared intention and modest sociality, where the conceptual, metaphysical, and normative resources at work in articulating those conditions come from the domain of individual

[173] An Essay Concerning Human Understanding second edition, Book II, chap. XXVII, paragraph 11.
[174] "Self to Self," 172. In Chapter 3 I cautioned against the overextension of this idea to all cases of intending, but this is not to reject the importance of this idea to a Lockean view of personal identity.
[175] "Self to Self," 192

planning agency. And, with one possible qualification, that is what the basic thesis tries to do. The possible qualification is due to the appeal within the basic thesis to an as-yet not-fully-analyzed notion of common knowledge. Nevertheless, the basic thesis, if true, would challenge claims – claims that are, in different ways, made both by Gilbert and by Searle -- that there is a fundamental non-reducibility, both conceptual and metaphysical (and, in Gilbert's case, normative), of shared intentionality, an irreducibility that goes beyond issues about the nature of common knowledge.

The basic thesis depends on the planning theory of individual human agency. The claim is that this planning theory, in going beyond the desire-belief model of individual agency, provides a model of individual agency that is sufficiently rich – conceptually, metaphysically, and normatively -- for use as a central element in a conservative construction of shared intention and shared agency. This is an aspect of the fecundity of planning structures.

Recall the observation that mutual, strategic adjustment that is in equilibrium within common knowledge – as in the case of walking alongside a stranger -- does not ensure modest sociality. As I have noted, this does not mean that the very idea of a kind of equilibrium within common knowledge is not important. Indeed, what the basic thesis provides is a model of shared intention as a complex kind of shared practical settledness within common knowledge. This is a shared practical settledness that involves planning attitudes, interlocking intentions, commitments to mesh, and so on – attitudes that engage distinctive norms of intention rationality.

As noted, the claim is not that the step from individual planning agency to modest sociality is simple or undemanding. Indeed, the basic thesis seeks to articulate the complexity and psychological richness of this step. You could be a planning agent and yet still not be capable of these further complexities. The basic thesis offers an articulated model of what is involved, at least in a central case, in being not only a planning agent but also a participant in modest sociality. And according to this model, the further developments of the psychic economy of planning agency that provide the bridge to modest sociality are themselves applications of conceptual, metaphysical, and

normative resources already available within the theory of individual planning agency.[176] This is the conservatism of our construction of modest sociality.

4. Social networks

The basic thesis highlights referentially interlocking and interdependent intentions of the different participants in modest sociality, intentions with a common social content. In this way the basic thesis sees the social glue characteristic of modest sociality as quasi-Lockean, since these inter-personal ties are to some extent analogous to the cross-temporal Lockean ties characteristic of individual planning agency over time.

This gives us resources to model a phenomenon that lies in the territory between a mere concatenation of different instances of modest sociality on the part of a range of different small groups, on the one hand, and on the other hand the shared intention and modest sociality of a larger group that consists of the members of those small groups.

Consider a structure of overlapping strands of the quasi-Lockean ties involved in instances of small-scale modest sociality. A and B, let us say, are engaged in shared activity J_1, where this involves their shared intention that they perform J_1. So there are ties of interlocking and interdependent intentions, with common social contents, between A and B. B and C, let us suppose, are engaged in a shared activity J_2, which involves an associated shared intention in favor of J_2. So there are ties of interlocking and interdependent intentions, with common social contents, between B and C. C and D, let us suppose, are engaged in J_3 … And so on, all the way to Y and Z. There is a

[176] Again, putting aside the exact status of a common knowledge condition.

We can see an evolutionary proposal of Michael Tomasello and Malinda Carpenter as an analogue of this proposal about the conceptual, metaphysical, and normative continuities in the step from planning agency to modest sociality. Tomasello and Carpenter write:

> The emergence of these skills and motives for shared intentionality during human evolution did not create totally new cognitive skills. Rather, what it did was to take existing skills…and transform them into their collectively based counterparts of joint attention, cooperative communication, collaborative action … Shared intentionality is a small psychological difference that made a huge difference in human evolution… (Tomasello and Carpenter, "Shared intentionality," Developmental Science 10:1 (2007), 121-125, at 124)

This evolutionary proposal has a similar structure to my conservative constructivism, though I am not making an evolutionary claim, and though Tomasello and Carpenter speak of various "existing skills" where I speak primarily of individual planning agency. And I share with Tomasello and Carpenter the idea that the "small psychological difference" involved in the step to modest sociality makes a "huge difference". While I do not see my conservative constructivism as depending on the truth of some such evolutionary story, this constructivism might be of use in articulating and assessing that evolutionary story by providing a model of what more precisely is involved in the cited transformation to "collectively based counterparts".

chain of overlapping quasi-Lockean inter-personal strands connecting A with Z, even if there are no direct ties of interlocking and interdependent intentions between A and Z. The relation between A and Z is in this respect parallel to the relation between the child and the senile general in Anthony Quinton's defense of the Lockean view of personal identity in response to Reid's famous example.[177]

There can be such overlapping strands of modest sociality in the absence of a shared intention on the part of the overall group. And many times the different J_i's in such overlapping strands will have only an incidental relation to each other; though we can nevertheless investigate how such overlapping social strands might propagate various kinds of social influences. (Certain kinds of crowds might be like this.) But there will also be cases in which the J_i's are related in ways that are theoretically important. Perhaps each of many overlapping pairs of agents is a lobbying group trying to exert influence on a specific issue in American politics, though there is no overall lobbying issue on which all are working.[178] Or perhaps each of many overlapping pairs of agents is a small business group that is trying to maximize the profit of that very group, though no one is trying to maximize the overall profit of the overall group (though perhaps each expects that to happen as a result of what each small group is doing).[179]

So consider a structure of overlapping strands of modest sociality where the act-types that are the target of the involved shared intentions on the part of the relevant small groups stand in an appropriate (so far, unspecified) relation. Using a currently popular way of talking, we might call this phenomenon a social network. Assuming the involved joint act types are indeed appropriately related, the group consisting of the members of all of the cited overlapping small-scale lobbying groups, or of all the cited overlapping small business groups, is, in this sense, a social network.

There may be no overarching shared intention in such social networks. And an absence of an overarching shared intention will induce important limits given the central role of shared intentions in framing bargaining and, as we will see, shared deliberation.

---

[177] Anthony Quinton, "The Soul," The Journal of Philosophy 59 (1962), section 2. Related and more complex ideas are in H.P. Grice, "Personal Identity," Mind 50 (1941). A more extended treatment of social networks would exploit the further complexities of inter-relation highlighted by Grice in this essay.
[178] I assume that a massive disjunction of issues is not itself an issue in the relevant sense.
[179] A variation on an example from John Searle, discussed above in Chapter 00. See his "Collective Intentions and Actions," at 404-405.

Nevertheless, a social network in this sense involves social interconnections that go beyond those involved in a mere concatenation of multiple instances of modest sociality on the part of different small groups.

Another kind of example involves overlapping strands of modest sociality over time. On natural assumptions, this is what happened in – to use an example, from Seamus Miller -- the construction of a certain cathedral over several hundred years. The various agents in the construction of the cathedral over those many years participated in a kind of social network.[180]

I will not try to say generally what relation is needed between the $\underline{J}_j$s for there to be a social network. Different ways of specifying the needed relations will give us different ideas of a social network, and it will depend on our theoretical interests how exactly we will want further to specify this idea. The point here is only that the basic thesis gives us resources to develop theories about such social networks of overlapping interconnections across multiple, local instances of modest sociality. It gives us these resources while highlighting both that such cases need bring with them no overall shared intention on the part of the overall group, and that the absence of such a shared intention will involve important limitations on shared reasoning.[181]

5. Treating as a means?

According to the basic thesis, when you and I share an intention to $\underline{J}$ I intend that we $\underline{J}$ in part by way of your intention that we $\underline{J}$, and vice versa. This semantic interlocking of our intentions is, according to the theory, an important aspect of the interconnections between us that are characteristic of shared intentionality. But, as Christine Korsgaard has emphasized (in correspondence and conversation), this can seem puzzling. In intending that your intention be effective in the pursuit of our joint

[180] Seamus Miller, Social Action p. 13. Miller sees such cases of "intergenerational joint projects" as posing a problem for a theory like mine; but I think that a treatment of such cases by way of the idea of a social network seems promising.

But what if the building of the cathedral comes to a complete stop for several generations, and then the project is re-discovered and people begin it anew and complete the cathedral? This kind of case that might motivate an extension of the idea of a social network, one that allows that some of the "links" across time may not themselves be ones of modest sociality.

[181] I do not claim that my theory is uniquely positioned to do this. For example, and as Luca Ferrero has observed, we could consider an extension of Gilbert's theory that focuses on overlapping strands of joint commitment. My concern here is only to indicate how my theory would treat these matters, not to argue that these are matters that could not be theorized within other frameworks.

activity, I may seem to be seeing your intention and your agency as, at bottom, a means to what I intend – namely, our joint activity. But is it plausible that at the heart of shared agency is an intention to treat the other as a means?[182]

Well, in intending that we J in part by way of your intention that we J, I am indeed intending that your intention play its role in guiding your activity as part of our activity. But recall that my intention not only interlocks with yours; it is also reflexive. I intend also that my own intention function in the standard way. So I am not only intending that your intention function in the connection between your thought and your and our action; I am also intending that my intention function in the connection between my thought and my and our action. In this sense my intention supports the agency of each of us. In so intending, then, if there is a sense in which I treat you as a means, it is a sense in which, given its reflexivity, my intention involves treating myself as a means. Since treating myself as a means in this special sense is, I take it, compatible with full-blown individual intentional agency, I infer that the sense in which I also treat you as a means (if there is such a sense) is compatible with full-blown shared intentional agency.

6. Deception and coercion re-visited

We can now return to the issue, noted in chapter 1, of the impact of deception or coercion on the status of our activity. And my proposal will be that the basic thesis supports a plausible picture of the complex interaction between coercion, deception, shared intentionality, and cooperation.

The first point is that the basic thesis helps us see the ways in which many forms of deception or coercion involve attitudes that baffle shared intentionality. If you are deceiving me about the finances of our house painting, despite my expressed intention that we stay within a certain budget, then you are thereby failing to be committed to our acting by way of sub-plans that mesh. After all, you are willing for us to paint together in a way that, as you know, violates my sub-plan that we stay within the budget. Further, we can suppose that I intend that in pursuing the joint activity I act on beliefs that are accurate.[183] And you are not committed to mesh with that part of my sub-plan. Again,

[182] I am indebted to Korsgaard for articulating this objection, and to Abraham Sesshu Roth for helping me answer it.

[183] Cp. Bernard Williams' appeal to our "general interest in being factually and rationally correctly informed," in his "Internal Reasons and the Obscurity of Blame," in his Making Sense of Humanity, at 37.

when you threaten me with your gun you my well be failing to intend that we paint by way of my intention that we paint. What you are doing may well be, rather, trying to insure that we paint by way of my intention to avoid your threatened violence.

More generally, it seems that many kinds of deception or coercion will involve attitudes that straightforwardly violate one of the main conditions cited by the basic thesis as central to shared intentionality. Deception will normally involve a disregard for meshing with the other's sub-plans, sub-plans that will normally include an intention to act on accurate beliefs; and coercion will normally involve a lack of concern with whether the agent is acting on an intention of his in favor of, in particular, the shared activity, or with whether that agent's sub-plans themselves include a rejection of coercion. And in violating these conditions cited by the basic thesis, these forms of deception and coercion will tend to undermine relevant inter-personal coordination of action and planning and the structuring of relevant bargaining and shared deliberation, that are the characteristic roles of shared intention.

The second point is that the basic thesis also helps us see how in some cases deception or coercion may fail to block shared intentionality, since they may fail to block one of these central conditions. Suppose you deceive me about your reasons for participating in our painting. But suppose, as you know, I don't care why you are participating: nothing about this gets into my relevant sub-plans. And suppose that if I did care about this, in a way that was internalized into my relevant sub-plans, you would relent in some way. If this were true then it may be that, despite your deception, you are in fact committed to our painting by way of sub-plans that mesh. Again, the master's power over the slave might perhaps be, so to speak, in the background. Despite this power, the master still intends that their sub-plans for picking the cotton mesh, and that the picking goes by way of each of their intentions in favor of the picking. And so there might be subtle forms of inter-personal coordination, and bargaining about means and preliminary steps – forms of coordination and bargaining that are characteristic of modest sociality.

These observations about these last two cases are compatible with the thought that deception and coercion quite generally block cooperative interaction --- where we are understanding cooperation in a way that is, in the ways discussed earlier, to some

extent moralized. In this sense, even in these latter two cases the agents are not engaged in a shared cooperative activity. But it seems that they may be engaged in shared intentional activity. This is the possibility, anticipated in Chapter 1, of shared intentional activity that is not shared cooperative activity. And the basic thesis helps us understand central ways in which deception or coercion block, and ways in which they may not block, shared intentionality. Deception and coercion block shared intentionality of the sort highlighted by the basic thesis, when they do block it, not primarily because they violate some moral ideal of human interaction (though they do violate such an ideal), but because they involve attitudes that violate the specific social-psychological conditions highlighted by that thesis. And these conditions are highlighted by the basic thesis not primarily because of some background moral ideal of human interaction, though a defender of the basic thesis may, of course, also defend some such ideal. These conditions are highlighted, rather, because they are elements of a social-psychological economy that – or so it is claimed – realizes the main social psychological features of modest sociality. They are features of a structure of inter-connected attitudes of each which is such that when it functions properly there will be the forms of social functioning and social rationality that are characteristic of modest sociality.[184]

7. The compressed basic thesis

Let us now return to the observation that condition **(ii)**, suitably understood, entails conditions **(i)** and **(iii)**. I have allowed these redundancies to stand in order to get as clear a view as possible of the relevant building blocks. But now that all the pieces of our construction are available and have been explained, it will be useful to formulate a compressed version of the basic thesis, one that avoids these redundancies. We can say that basic sufficient conditions for our shared intention to J are as follows:

A. Intention condition: We each have interlocking and reflexive non-attenuated intentions in favor of our J-ing by way of each of these intentions and mutual responsiveness in sub-plan and action, and so by way of sub-plans that mesh.

[184] Aspects of this section derive from thoughts triggered by several discussions with Christine Korsgaard.

B. Belief condition: We each believe that the intentions of each in favor of our J-ing will be jointly effective, in part by way of relevant mutual responsiveness; and we each believe that there is interdependence in persistence of those intentions of each in favor of our J-ing.

C. Interdependence condition: There is interdependence in persistence of the intentions in favor of our J-ing.

D. Common knowledge condition: It is common knowledge that **A.-D**.

And we can say that what further is needed for shared intentional activity, and so for modest sociality, is that this shared intention to J lead to our J-ing in accordance with the following connection condition:

E. Mutual responsiveness condition: our shared intention to J leads to our J-ing by way of mutual responsiveness in sub-intention and action that tracks the intended joint activity, in conditions of common knowledge.

The compressed basic thesis is the claim that conditions **A.-E**. provide sufficient conditions for shared intention and modest sociality. And this compressed basic thesis supports the continuity thesis.

At the heart of the theory, then, is a quintet of conditions: the participants have interlocking and reflexive intentions in favor of the joint activity by way of mutual responsiveness and so mesh; the participants have relevant beliefs about efficacy and interdependence; there is interdependence in persistence of relevant intention; there is common knowledge; and there is mutual responsiveness. When we step back to see the basic outlines of the theory it will be this quintet that will be in view, though in many cases it will be important to keep track of details that are more easily accessed by appeal to the non-compressed version of the thesis.

8. Too demanding?

The compressed basic thesis can help us think about the worry that our model of modest sociality is too psychologically demanding. We can see this as a worry addressed primarily to the intention condition. This condition requires intentions that are interlocking, reflexive, and favor the joint J-ing by way of mutual responsiveness, and so by way of meshing sub-plans. But, it is natural to object, couldn't there be agents --

four-year old humans, perhaps -- who engage in a form of modest sociality but for whom such complexity is not yet psychologically available?

Well, the level of psychological complexity available to a four-year-old human is a complex empirical issue. But however this empirical issue turns out, we need to understand this concern within the context of two important points. The first is that the complex content of the intentions cited in the intention condition may be only implicit. The agents need to have relevant intentions and associated dispositions of tracking, adjustment, and responsiveness; and this web of intentions and dispositions needs to support the attribution of intentions with the complex content cited in the intention condition. But this may not involve explicit conscious awareness of this complex content. So the intentions cited in the intention condition are not as psychologically demanding as they may at first seem.

The second point concerns the central aim of the (compressed) basic thesis. This thesis aims to show how a robust form of sociality – as I have called it, modest sociality -- can be constituted by structures that have their home in the planning theory of individual agency; and it aims to show that the distinctive normativity of modest sociality will emerge from those individualistic structures. In this way it aims to show that there need not be a deep conceptual or metaphysical or normative discontinuity between individual planning agency and modest sociality. That is why its fundamental concern is with sufficient conditions for modest sociality.

If the basic thesis succeeds in these efforts it will thereby have defended the continuity thesis. We would then be at liberty to go on to investigate whether certain less conceptually sophisticated psychological phenomena might in certain circumstances substitute for these more conceptually demanding attitudes of each. This might happen in tandem with, and be shaped by, the kinds of planning structures emphasized by the basic thesis. When we dance together, for example, our shared intention to dance shapes our behavior by way of complex forms of interpersonal bodily responsiveness that may not be included in the contents of our intentions.[185] Again, such conceptually less demanding mechanisms might turn out to be common in the

[185] A point emphasized in conversation and correspondence by Alison Gopnik

sociality of younger human children[186] -- though (as Christopher Kutz has emphasized in conversation) we also need to keep a lively sense of the differences between the shared activities of 4-year old humans and the swarming of bees. We are also at liberty to allow that the attitudes cited by the basic thesis need not be the actual causes of relevant behavior but function only as what Philip Pettit calls "standby factors" that exercise "virtual" rather than "active" control.[187] And all this is compatible with the claim that the (compressed) basic thesis articulates the structure of a fundamental form of modest sociality, and that this structure is continuous – conceptually, metaphysically, and normatively -- with those at work in individual planning agency.

[186] For a view like this see Deborah Perron Tollefsen, "Let's Pretend! Joint Action and Young Children," Philosophy of the Social Sciences 35 (2005): 75-97. For a related but different idea see the explanation of joint action in Stephen Butterfill, "Joint Action and Development".
[187] Philip Pettit, A Theory of Freedom (Cambridge, England: Polity Press, 2001) at 38-39.

Chapter Five: Modest Sociality and Strategic Interaction

A central thought of this discussion is that modest sociality, while consisting in appropriate forms of inter-connected planning agency, is not merely strategic interaction within a context of common knowledge. As we noted in Chapter 1, when two strangers walk near each other at a similar pace and in the same direction down Fifth Avenue, and do so without collision, each might be acting strategically in the light of his knowledge of what the other does. Each might be seeking to achieve what he wants or values in light of his prediction of how the other will act, knowing full well that the other is reasoning in a parallel fashion. In this sense their patterns of interaction might be in strategic equilibrium in a context of common knowledge. Yet they still might not be engaged in modest sociality: they still might not be walking together in the sense we are after. An adequate theory needs to explain what is central to cases of modest sociality that is missing from such cases of merely strategic interaction.

My aim in this chapter is to explain how the basic thesis satisfies this condition of adequacy. To do this we will need to return to the point that the intentions that are at the heart of the basic thesis are to be distinguished both from expectations and from evaluations. And we will also need to highlight the way in which the basic thesis appeals to inter-connections across the participants that go beyond the epistemic inter-connection of common knowledge.

1. Modest sociality, intention and expectation

Return to the case of your walking alongside a stranger. As noted, you and the Stranger may be acting in ways that are in strategic equilibrium in a context of common knowledge. Each knows both what the other intends and does and that each acts because of his intention so to act; each acts in the light of this knowledge of the other; and each knows that if both so act there will be a coordinated concatenation of their walking actions. However, this case need not satisfy many of the conditions central to the basic thesis.

First, though each believes that there will be the cited coordinated concatenation of walking actions it does not follow that each intends that. To intend the coordinated concatenation each would need to be disposed to take it as an end for his own means-

end reasoning and to be guided in action by this end. And to intend the coordinated concatenation each would need to be disposed to filter potential options for deliberation with an eye to their compatibility with this end. But it may be that none of this is true of you and the Stranger, even though your actions are in strategic equilibrium within common knowledge. Perhaps, for example, the Stranger is looking for ways to thwart your progress down the street without physically interfering with you, even though he sees that you are indeed progressing down the street and he is doing what he thinks best given that your are. This Stranger does not intend that the two of you walk together down the street, or that your sub-plans mesh. Given what he knows to be the limits on his powers, he does expect that you will in fact walk in the way you are walking. And he intends to respond to that, and so expects that there will in fact be a coordinated concatenation of the walking actions of each. But this is not yet to intend that coordinated concatenation.

Nor need this Stranger be disposed to help you if you need it. And his responsiveness to you may not track simply the coordinated concatenation but rather a complex end that also includes non-violently thwarting you if, contrary to his expectation, it turns out that he can.

Again, perhaps the Stranger expects that your walking will be the issue of your relevant intentions and yet does not intend that. Perhaps he prefers a mechanism that would issue in your walking in a way that bypasses your intentions – perhaps, as in a mafia case, by pushing you along.[188] Being a realist and not especially strong, however, he does not believe that this is what will happen; so he expects that you will walk by way of your relevant intention, and he does what he sees as best given that. But he does not intend that your intention be efficacious. So his intentions do not appropriately interlock with yours.

It follows that this case of walking alongside a stranger does not satisfy the conditions set out in the basic thesis. So the basic thesis can say that this is a case of strategic interaction that is not a case of modest sociality.

[188] This contrasts with a case in which the Stranger has no preference as between these different mechanisms. As I indicated in Chapter 2, in discussing Miller's example of the tunnel builders, in the absence of some such preference he might still intend that your actual intention be efficacious.

This result depends on taking seriously the idea that the basic thesis requires that what is intended (and not merely expected) by each participant includes the joint action, where that includes the other's role in it. If all that were required were that each expects the other to intend and act in relevant ways and then intends his own actions in the light of that expectation, the account would be too weak: we would not thereby have ensured relevant dispositions to track and not to thwart the other's role in the joint activity, to filter options with respect to that activity, to help the other if need be, and so on.[189] But it is fundamental that the basic thesis requires that each intends the joint action, where this joint action includes the roles of each. And intention differs systematically from ordinary expectation.

Consider now a closely related objection from Bjorn Petersson, who writes:

> Suppose I want the window smashed. When I note your presence on the street, I think that if you act in a certain way, the window can be smashed as a result of both our acts, and I form an intention accordingly. What I intend in that case is merely to get the window smashed, while predicting that your actions will be components in the process leading to that result. This prediction may rest upon my knowledge that your intentions are similar to mine, and that our subplans are likely to mesh in a way that enables me to reach my goal. There is mutuality and interdependence, in line with Bratman's requirements. Still, I would say, nothing in this picture captures "sharedness" or "collectivity" in any sense distinct from what we can construe in terms of standard individualistic theory of action.[190]

[189] This seems to be a problem for Nicholas Bardsley's positive proposal in his "On Collective Intentions: Collective action in Economics and Philosophy." See his (2) on p. 152. Bardsley would respond that on his account what each expects is not just the actions of the others "but their disposition to act on team considerations."(156) Without pausing to examine the details of this cited disposition, we can make the point that if all that we say is that each expects this, then we have not yet ensured that each intends to support it. I might, for example, believe you are disposed to act on such considerations but not be set to filter options of mine that are inconsistent with your so acting. The same point can be made in response to Bardsley's condition that each participant expects the other's actions "to flow from an intention like their own".(158) I can expect that but still aim to prevent it. The lesson is that expectation is not a sufficiently practical commitment to constitute shared intention.

[190] "Collectivity and Circularity," 140-141. Petersson refers to my "Shared Cooperative Activity," and my "I Intend that We J".

Petersson thinks that a theory like mine will be led to say that the case he describes is a case of "sharedness"; and Petersson thinks (correctly, in my view) that this would be a mistake. What to say?

An initial point is that this case, as described, seems so far not to satisfy the conditions cited in the basic thesis. This is in part because Petersson's description appeals at crucial moments to expectation when what is required by the basic thesis is intention; and it is a central theme of the planning theory that these attitudes differ in systematic ways.[191] Though in Petersson's example I expect that you will act in ways that promote the smashing of the window, it is not clear from the description of the example that I intend not just my own act of window smashing but a window smashing that involves both of us as agents. It is not clear that I intend that we smash the window. So it is not clear that I am set to be relevantly responsive to you in intention and action. It is not clear, for example, that I am set to help you if needed.[192] Nor is it clear in Petersson's example that I intend that our sub-plans mesh, though I do expect that they will. So in the case as described by Petersson I may not satisfy my side of a shared intention, as that is understood by the basic thesis. Further, Petersson says nothing about how the intentions of the other person connect to mine, though this too will be essential to shared intention. After all, it is fundamental to the basic thesis that shared intention involves both of us: it is not just a matter of what is in my head. Nor is it part of Petersson's case that there is joint-action-tracking mutual responsiveness in intention and action, or that all this is common knowledge. So I agree with Peterson that there is neither shared intention nor "sharedness" nor modest sociality in his case, at least as his case is naturally understood; but I do not see that this is yet an objection to the basic thesis.

[191] For what may be a similar mistake see Michael Bacharach (edited by Natalie Gold and Robert Sugden), Beyond Individual Choice (Princeton University Press, 2006), at p. 139. As Marc Pauly once remarked, it may be that this mistake is rooted in a failure fully to appreciate the significance of the introduction of distinctive planning structures, over and above desire-like and belief-like structures, within the theory of individual agency.

[192] Though, in fairness to Petersson, in the earlier essays to which he refers this condition of being set to help if needed is cited as a condition of shared cooperative activity, rather than a condition on shared intention. See "Shared Cooperative Activity," in Faces of Intention at 103-4. Still, the other features I cite as elements of shared intention are all built into the earlier discussion as well as this present discussion.

Nevertheless, Petersson may think that even once it is made explicit that these further conditions of the basic thesis are satisfied by (an upgraded version of) his example, we will still not have "sharedness". But at this point much of our discussion in Chapters 2-4 comes into play. Conformity to the basic thesis – and the distinctive forms of interrelation across intentions that it highlights -- ensures both standard forms of functioning of shared intention in linking us together and organizing our thought and action, and associated forms of social rationality. There is here, in the satisfaction of the conditions cited in the basic thesis, a significant social glue.

Does this suffice for "sharedness"? Perhaps we can capture one of the intuitions behind Petersson's concern by returning to the distinction, from Chapter 3, between reason-dependence and mere enabling dependence. In a kind of case toward which Petersson points each treats the other's intention opportunistically as an enabling condition of the joint activity, and not as a reason for favoring the joint activity. Nevertheless, for this to pose a problem for the basic thesis we must go on to suppose that the various conditions of that thesis are satisfied: each intends that things proceed by way of each other's intentions, mutual responsiveness, and so meshing sub-plans; each is in fact appropriately responsive to the other; each is at least minimally disposed to help the other if needed; each is set not to thwart the other; there is interdependence in intention; the participants' intentions are semantically inter-connected; there is, as a result of all this, rational pressure in the direction of social rationality; and each expects all of this to issue in the joint action each intends. Once this is all made explicit it seems to me plausible to say that this is a version of the kind of "sharedness" in which we are interested, though I grant that such a case may differ in certain ways from the kind of reason-for interdependency characteristic of Romeo and Juliet (as commonly understood).

Natalie Gold and Robert Sugden explore a related worry. They consider a Hawk-Dove game characterized by the following pay-off matrix for two agents:[193]

	dove	hawk
dove	2,2	0,3

[193] Natalie Gold and Robert Sugden, "Collective Intentions and Team Agency" Journal of Philosophy 104 (2007): 109-137, at 112.

hawk	3,0	-5,-5

Here the first number in each box represents the "payoff" for the first player, the player who is choosing between the two different rows; the second number represents the payoff for the second player, the player who is choosing between the two different columns. In two plays of this game – namely, (dove, hawk) and (hawk, dove) -- each agent's response is the best response (as measured by his payoff) to the other's response. Neither agent can improve his payoff by unilaterally changing what he chooses, holding fixed what the other chooses. Each case involves, in this sense, a Nash equilibrium. Suppose then that the players arrive at (dove, hawk) in a context of common knowledge. Must the basic thesis say that there is then a shared intention in favor of (dove, hawk)? If so, the basic thesis will be too weak, for it will have failed to "differentiate collective intentions from the mutually consistent individual intentions that lie behind Nash equilibrium behavior in games."[194] What to say?

Well, when these players arrive at (dove, hawk) in a context of common knowledge each knows what the other intends to do and does; each knows that if each does as he intends the upshot is (dove, hawk); and each knows that this outcome is a Nash equilibrium. But it does not follow that each intends that (dove, hawk) be the upshot, or that each intends that the other's intention be successfully executed. After all, to know need not be to intend. Nor does it follow that there is any intention on the part of each that, were there to be a need for sub-plans (a potential need that is not represented in the cited pay-off matrix), these sub-plans are to mesh. Nor is there any guarantee of any disposition on the part of each to help the other if need be. Further, there is no guarantee that each will be committed to filter out actions that would thwart the other. On a natural understanding of the case, were the first player in the (dove,

[194] "Collective Intentions and Team Agency," p. 109. In putting the point this way Gold and Sugden are glossing over the distinction between intentions of individuals that are essential to shared collective activity, and collective intentions in the sense of intentions of a collective. The latter would be an analogue of my talk of shared intention, the former an analogue of Searle's talk of we-intentions in the head of individuals. Given that they see collective intentions as the output of team reasoning, and (as we will see below) team reasoning as the reasoning of an individual, it seems that it is the former Searle-type thought that they have in mind here. But the latter reading is needed for their discussion to provide the intended criticism of my view. So I will take it that their objection to the basic thesis is that it does not distinguish "intentions that lie behind Nash equilibrium behavior" from shared intention.

hawk) scenario to see a way of tricking the other into playing "dove" he would. Nor is there a guarantee of any disposition on the part of each to respond to the other in ways that track the supposedly joint action of (dove, hawk) rather than an alternative that would be higher ranked by that individual.

As in the case of Petersson's objection, then, a clear view of the difference between expecting the other to intend and act in a certain way, and intending both the joint action and that the other's relevant intention be successful – and related phenomena of intending responsiveness and mesh, of willingness to help if needed, and of actual mutual responsiveness in intention and action that tracks the joint outcome – blocks the present worry that the basic thesis does not adequately capture "collective" or shared intentions. So I think that Gold and Sugden may not be sufficiently attentive to, in particular, the difference between intending and believing.[195]

So we have a trio of cases of strategic equilibrium within common knowledge: walking alongside a stranger, the initial version of Petersson's window-smashing case, and Gold and Sugden's hawk-dove case. In each case we have seen how a defender of the basic thesis can plausibly claim that the conditions cited by that thesis are not satisfied. And in each case this reply depends on taking seriously the idea that the basic thesis requires that what is intended (and not merely expected) by each participant includes the joint action (where that includes the other's role in it).

To be sure, in each of these cases of strategic interaction, the participants intend to act in certain ways given that, as they expect, the other will act in certain ways. But in each case we should resist the inference from S intends A, given that (as she expects) B, to S intends (A&B). After all, S may intend A, given that (as she expects) B without any disposition at all to try to ensure or promote B, or to filter out options incompatible with B. And when we resist this inference, and insist on the distinction between intending and expecting, we are in a position to, in the words of Gold and Sugden, "differentiate collective intentions from the mutually consistent individual intentions that lie behind Nash equilibrium behavior."

[195] A sign of this is their remark that "Bratman might object that it is too glib to interpret P1's intending that J come about 'because of' P2's intention as the idea that P1 believes that P2 has the corresponding intention and acts on the basis of this belief. But his expansion of it … is opaque."(115) On my view, as I have indicated, this proposed identification of intending with belief is a mistake.

2. Modest sociality, intention and evaluation

So, one key to seeing how the basic thesis treats the distinction between modest sociality and mere strategic interaction is to keep track of the distinction between intending and expecting. I have also noted reasons for keeping track of the distinction between intending X and judging X best, or ranking X highest in one's relevant evaluative ranking. An individual might intend X even while thinking some alternative is as good or is in a relevant way incomparable, or while simply being unsure which alternative is best. And an individual might even, in a case of weakness of will, intend X while judging that some alternative is strictly better. And similar points hold for shared intention. We can share an intention to X even if none of us thinks X is strictly superior to its alternatives. And we can have weak-willed shared intentions in which we each see some alternative as better. With this second aspect of the distinctiveness of intention in mind, let's return to Gold and Sugden's essay.

Begin by noting that Gold and Sugden think that my appeal to mutual responsiveness does to some extent help distinguish collective intention from mere Nash equilibrium. But they think it does not go far enough:

> To illustrate mutual responsiveness in action, Bratman gives the example of singing a duet, each singer accommodating herself to the actions of the other. The intuition is clear enough, but there is no analysis of how mutual responsiveness works.[196]

It is clear that the mutual responsiveness Gold and Sugden have in mind here is, in particular, relativized and joint-action-tracking mutual responsiveness. And they think that such mutual responsiveness <u>is</u> part of a solution to their problem. But they also think – correctly -- that there are further theoretical issues we need to solve in order fully to understand such mutual responsiveness. And Gold and Sugden see themselves as providing a theory of how this mutual responsiveness works. This is their theory of "team reasoning". In such reasoning, roughly, each reasons, in a context of common knowledge, from her assessment of what it would be best for a certain group or team to do, an assessment on which the participants converge.[197] Each reasons from this

[196] "Collective Intentions and Team Agency," p. 116.
[197] 124-5

assessment, together with her "identification" with that group or team, to an intention to play her role in that supposed-best team option.[198] "Collective intentions" are the outputs of such team reasoning.

Note that "team reasoning" is not reasoning that is engaged in by a team.[199] Team reasoning is reasoning of an individual; but it is reasoning of an individual in which the evaluative ranking that the individual brings to bear captures how that individual ranks options from, as it were, the point of view of the team.

Though it is the individual who is engaged in this team reasoning, Gold and Sugden assume that all the different members of the relevant "team" have the same team-ordering of the different options open to the team. More precisely, the supposed common ordering is represented as each wanting "the value of U to be maximized" where U is a "payoff function" that is common to all the members and is supposed to capture an ordering from the point of view of the group.[200]

This is to see a shared intention in favor of X as involving an evaluative "team" ranking that is common to all the members of the "team," and in which X is at the top. But, as Randall Harp has emphasized, the idea that in shared intention the parties converge on the same evaluative ordering from the team's perspective is in tension with the distinction between intention and evaluation.[201] It seems that we can have a shared intention to go to concert C even if neither of us thinks that is the best option from the point of view of the group. (Perhaps you think concert A is best, and I think concert B is best. Or we are each simply unsure which option is best.) So it seems that there can be shared intention in the absence of agreement on the best, contrary to what is assumed by Gold and Sugden. And once such a shared intention is in place, it is that shared intention, in contrast with a common team evaluative ordering, that structures bargaining and shared deliberation aimed at achieving relevant mesh in sub-plans.

---

[198] For this talk of identification see their schemas 3 and 4, pp. 125-6.

[199] Gold and Sugden say that they are interested in "modes of team reasoning that are used by individuals as members of teams." (117) Indeed, if the theory were depending on an appeal to the shared activity of the team in reasoning together it would be making use of the very idea – that of a shared intentional activity – that we are trying to understand.

[200] 124-5

[201] See Randall Harp, Collective Goals, Collective Reasoning, Collective Action (Ph.D. Thesis, Stanford University 2009) chap. 3 section 2.

There will also be problems with the model even in certain cases of evaluative convergence. First, it might be that the participants agree in their team evaluative ranking, but that these agreed rankings do not specifically support the shared intention. Perhaps they agree that X and Y are incomparable, and yet still have a shared intention in particular to X. Second, perhaps the participants agree that X is best and yet, out of weakness of will, have instead a shared intention to Y. Two weak-willed lovers might engage in relevant shared activities all the while agreeing in their judgments that an alternative would be best. So even if we assume that there is evaluative convergence across the participants, we cannot in general see shared intention as a matter of that evaluative convergence.

This is not to deny that there is a phenomenon of team reasoning of the sort that is described by Gold and Sugden, or that this phenomenon may many times be a sensible technology for participants in a shared intention to use to figure out what to do.[202] But once we have a sufficiently nuanced understanding of the relation between intention and evaluation, Gold and Sugden's model of team reasoning seem ill-suited to explain what such a shared intention is.

3. Modest sociality and inter-connected planning agency

According to the basic thesis the intentions of the participants in modest sociality are inter-connected in ways that go beyond common knowledge. Their intentions in favor of the joint activity interlock and favor mutual responsiveness and meshing sub-plans; and there is interdependence in the persistence of these intentions. These inter-connections occur within a context of common knowledge, but they go beyond common knowledge: they involve practical commitments on the part of each that depend on and are in support of the practical commitments of the others.

This leads to another contrast with the Gold-Sugden model. As we have seen, that model highlights a common "team pay-off function" in a context of common knowledge. On the theory, the team pay-off function of each participant stands in two important relations to the corresponding team pay-off function of the other participants. First, these pay-off functions match. Second, this is all in a context of common

[202] Though I argue in Chapter 8 that Gold-Sugden team reasoning should be distinguished from an important kind of shared deliberation.

knowledge. In the previous section I argued, in agreement with Harp, that the assumption of match was too strong to be required for shared intention: we could share intentions despite associated disagreements in pay-off functions. I also argued that the assumption of match, even within common knowledge, was too weak to ensure shared intention: there could be such match in the absence of a corresponding shared intention, as in certain cases of weakness of will. I now want to make the further point that the cited epistemic inter-connection of common knowledge is too weak to capture inter-relations that are important to modest sociality.

Suppose you and I both have a team pay-off function that ranks as best, from the point of view of the team, our going to NYC. And suppose this is common knowledge. To return to our earlier mafia example, it still may be true that each of us intends to by-pass the relevant intention of the other in the process that will lead to our going to NYC. Perhaps each intends to throw the other into the trunk of his car. There is a common team pay-off function that ranks highest our going to NYC, and there is common knowledge of this; but there is no shared intention to go to NYC.

One of the ideas of the basic thesis is that shared intention involves a rich structure of practical inter-connections between the intentions of each. These practical inter-connections go beyond the epistemic inter-connection of common knowledge of common evaluation. And these practical inter-connections are not present in the mafia case. In contrast, the Gold-Sugden theory limits its account of relevant inter-connections to common knowledge of common evaluation.

In short, we don't get shared intention in favor of our J-ing just by putting "we J" into the heads of each -- perhaps in terms of team preference rankings in favor of our J-ing -- and then adding a common knowledge condition. There also need to be practical inter-connections of the sort highlighted by the basic thesis.

Perhaps Gold and Sugden would reply by appealing to their idea that each participant identifies with the group.[203] Perhaps they would say that the gang members just considered do not identify with the group consisting of both of them. But what is it to identify with a group? Gold and Sugden explain this as follows: "i identifies with G if i conceives of G as a unit of agency, acting as a single entity in pursuit of some single

[203] See schemas 3 and 4, pp. 125-6.

objective."[204] The problem is that this talk of "conceives" is just too cognitive to capture what is needed here: after all, each of the gang members can conceive of the two of them acting as a unit of agency. The problem is that neither intends that there be group agency in a sense that involves the effectiveness of the corresponding intentions of the other. But such interlocking intentions would involve practical inter-connections that go beyond the epistemic inter-connection of common knowledge, and so go beyond the resources of Gold and Sugden's theory.[205]

4. Taking stock

I conclude that the basic thesis has sufficient resources to articulate fundamental differences between modest sociality and merely strategic interaction. Objections to the contrary from Petersson, and Gold and Sugden seem to me not to succeed, and to depend to some extent on a failure to appreciate the first aspect of the distinctiveness of intention, the distinction between intention and expectation. Further, when Gold and Sugden offer an alternative account – one that aims to do a better job of saying what

---

[204] 125

[205] Consider finally a Gold-Sugden proposal concerning the relation between our views:

> Bratman's analysis is at the strategic level. … in shared cooperative activity, each agent has the intention that "we" perform some joint activity through the meshing of "my" sub-plans with "yours". This intention is not linked to any particular combination of sub-plans; rather, it expresses a commitment to engage with the other in a process of "mutual responsiveness" and "mutual support" which is directed towards the meshing of sub-plans in general. … [This] analysis leaves open the question of how, at the tactical stage, the members of a group coordinate their actions so that together they achieve their joint objective.
>
> Since this is the central question addressed by the theory of team reasoning, the two approaches can be seen as complementary. (136-7)

This assumes that such tactical reasoning will be Gold-Sugden team reasoning and I have argued that that theory fails to do justice to the distinction between intention and evaluation. But in any case, note that this picture seems to withdraw the objection to my model of "strategic" shared intentions, an objection I discussed above in section 1. The point now is only that we need more theory to understand exactly how the mutual responsiveness cited by my theory works "at the tactical stage".

This apparent reconciliation is, however, in tension with the closing remark that "the key difference between the two kinds of intentions [that is, collective intentions and "the mutually consistent individual intentions that lie behind Nash equilibrium behavior"] is not a property of the intentions themselves, but of the modes of reasoning in which they are formed." (137) Given the proposal about how our two approaches fit together, it seems that at most this comment applies to the "tactical stage" in which, according to Gold and Sugden, each arrives at an intention concerning his own part in the shared activity by way of, in particular, "team reasoning". Shared "strategic" intentions provide part of the background for the kind of reasoning they cite, reasoning that leads to "tactical intentions".(136) There is, so far as I can see, no argument that such shared "strategic" intentions must themselves be the issue of prior "team reasoning". (Raimo Tuomela makes this point in his "Collective Intentions and Game Theory," The Journal of Philosophy 106 (2009): 292-300, at 299.) And we have seen reasons not to expect that whenever there is such a "strategic" shared intention there is the kind of evaluative agreement built into "team reasoning".

the difference is between modest sociality and merely strategic interaction – their account of team reasoning fails to do justice to the second aspect of the distinctiveness of intention, the distinction between intention and evaluation.[206] Finally, in contrast with the basic thesis, the Gold-Sugden model has a purely epistemic account of the interconnections that are at the heart of modest sociality; and I have argued that this is a weakness in that model.

The basic thesis distinguishes modest sociality from strategic interaction by appeals to distinctive planning attitudes and to certain interconnections between those attitudes. In characterizing these interconnections the basic thesis does not make an essential appeal to inter-relations of mutual obligation. But at this point it might be thought by some that there is, in fact, a deep connection between modest sociality and mutual obligation, and that this is at the heart of the distinction between modest sociality and mere strategic interaction. If one has this thought one will suspect that though the basic thesis is right in going beyond common knowledge, it still does not adequately characterize the inter-personal interconnections that are the heart of modest sociality. I turn next to this challenge.

[206] Petersson offers a very different positive proposal in light of his challenge to my theory. I discuss Petersson's proposal in Chapter 7.

Chapter Six: Modest Sociality, Mutual Obligation

It can seem that if you and I share an intention to paint the house together then there are distinctive, corresponding obligations of each to the other, obligations that include obligations to play one's part unless one has been given permission by the other to opt out. And it can seem that these obligations are in some way essential to the shared intention. As noted, this idea is central to the work of Margaret Gilbert.[207] And the purported objection is that the basic thesis does not adequately provide for this essential feature of shared intention.

In the background is the combined thought that (a) the participants in modest sociality are tied together in distinctive ways that go beyond common knowledge, and that (b) it is an essential feature of these ties that they involve distinctive obligations and entitlements. In response, I propose to accept (a) but reject (b). As I see it, the central explanatory roles of shared intention do not require appeal to such obligations; it is prima facie plausible that there can be shared intention in the absence of such obligations; and we should distinguish between shared intention and a structure of mutual obligations in part because the latter may not play relevant explanatory roles. This leads me to conclude that nothing about modest sociality supports the introduction of forms of mutual obligation that are not already introduced by familiar moral standards.

I'll begin by reflecting on how the basic thesis aims to understand the interconnections between the participants and the associated explanatory structures.

1. Social ties, explanation, and social rationality without mutual obligation

The basic thesis highlights social psychological structures of inter-connected attitudes of individual planning agents, and it sees these structures as helping to explain

[207] See for example her essays: "What Is It for Us to Intend?" in her Sociality and Responsibility (Lanham, MD: Rowman & Littlefield, 2000), 14-36; "A Theoretical Framework for the Understanding of Teams," in Natalie Gold, ed., Teamwork: Multi-disciplinary Perspectives (New York: Palgrave MacMillan, 2005), 22-32; "The Structure of the Social Atom: Joint Commitment as the Foundation of Human Social behavior," in Frederick F. Schmitt, ed., Socializing Metaphysics (Rowman & Littlefield, Inc, 2003): 39-64; and "Shared Intention and Personal Intentions," Philosophical Studies 144 (2009): 167-187 Abraham Sesshu Roth develops a different version of this idea in his "Shared Agency and Contralateral Commitments," The Philosophical Review (2004): 359-410. I will mostly concentrate on Gilbert's more developed theory, but I will return briefly to Roth's alternative proposal below in note 00.

As noted earlier, there are also related ideas in Stephen Darwall, The Second-Person Standpoint.

significant aspects of modest sociality. In particular, it aims to provide resources for explanations of the forms of functioning that are central to shared intention: inter-personally coordinated action and planning in the pursuit of a common end, and associated bargaining and shared deliberation. And it aims to provide resources for explanations of associated norms of social rationality.

Let's begin by reviewing and to some extent supplementing our discussion in Chapter 4 of how the basic thesis seeks to achieve these aims.

Suppose that our walking together satisfies the various conditions of the basic thesis. So I intend that we walk in part by way of your corresponding intention, mutual responsiveness, and so meshing sub-plans. This complex content of my intention connects it with your intentions and thereby imposes rational pressure on me, as time goes by, to fill in my sub-plans in ways that fit with and support yours as you fill in your sub-plans. This pressure derives from the rational demand on me to make my own plans coherent and consistent, taken together with the ways in which reference to your intentions enters into the content of my intentions. Rational pressures on me to be responsive to and to coordinate with you are built into my own plans, given their special content and given demands of consistency and coherence directly on my own plans (demands that are a part of the planning theory of individual agency). And similarly with you. Assuming that each of us is a planning agent who tends to be guided by such rational pressures, and given assumptions of common knowledge, we have an explanation of various forms of social responsiveness that are characteristic of modest sociality. We have, for example, an intention-based explanation of why each seeks to keep pace with the other. And if one or both of us fails to be appropriately responsive to the other, while continuing to be participants in the shared intention, we have an explanation of the sense in which there has thereby been a rational breakdown.

Again, the model provides an explanation of the way in which shared intentions frame bargaining about means and the like. In shared intention of the sort described by the basic thesis, we each intend the shared activity in part by way of the intentions of the other and by way of meshing sub-plans. So we are each under rational pressure to seek to ensure that our sub-plans, taken together, both are adequate to the shared task

and do indeed mesh. And that is why, in the absence so far of adequate and meshing sub-plans, our shared intention will tend rationally to motivate, to structure and to constrain bargaining in the pursuit of such sub-plans.

Further, the basic thesis helps explain a kind of rational stability of shared intention, given that the intentions of each of the participants are themselves subject to characteristic rational pressures for stability. This is in part because of the interdependence in persistence characteristic of shared intention. There will be what we called in Chapter 4 a mutual stabilization, given the interaction between rational pressures for stability of the intentions of each and the interdependence in persistence of those intentions.[208]

Finally, this model can provide for these explanations even if there are significant differences in each agent's reasons for participating in the sharing. Though we participate for different reasons, our shared intention nevertheless establishes a shared framework that can explain downstream thought and action.

These explanations suppose that the functioning of the planning agency of the individual participants will be guided by basic norms of individual intention rationality – where these includes norms of consistency, agglomeration, coherence, and stability. Failures of such guidance will be rational breakdowns. And associated norms of social rationality emerge from such individualistic rationality. The appeal is, at bottom, to the relevant, inter-related intentions of the participants in a context of common knowledge, and to relevant norms of individual intention rationality, not to obligations the participants have to each other. And the claim is that these inter-related intentions, in these contexts and guided by these norms of intention rationality, provide a basic structure for explaining the main contours of shared intentional activity, including coordinated action and planning in the pursuit of a common end, and associated bargaining and (though this is a matter I postpone until chapter 8) shared deliberation. These forms of explanation may well be supported and supplemented by appeal to various forms of mutual obligation, if such there be and if the parties care about their

[208] As noted, this is what Facundo Alonso calls a "virtuous circle of mutual reinforcement".

conformity with these obligations; but such mutual obligations are not an essential aspect of these basic explanatory structures.

Consider the practical reasoning of the participants in such a shared intention. If you and I share an intention to paint the house I can reason as follows (though this is of course highly stylized): "I intend this joint activity, and so do you; and we both intend that this proceed by way of mutual responsiveness and sub-plans that mesh. And these intentions of each of us are interdependent in their persistence. I can promote what I intend by settling on my buying the paint, given my belief that if I do that you will go ahead and bring the paint brushes and so our sub-plans will both be more or less adequate and mesh with each other. And if you were knowingly to fail to coordinate with me in this way I could point out to you that this would be out of sync with what you (and we) intend, and so a kind of rational breakdown on your part. So: let me buy the paint as my part in our joint project!" When you also reason in similar fashion, and thereby arrive at an intention to bring the brushes, we thereby proceed with our joint project. And in our practical reasoning neither of us needs to appeal to thoughts along the lines of: "you owe it to me to continue with our project and to mesh your plans with mine,"[209] or other thoughts about other kinds of distinctive moral significance of our sharing.[210]

Granted, we are agents who, as participants in modest sociality, do frequently assure each other, make agreements with each other, promise each other, intentionally induce or reinforce reliance on each other, [211] and so on. And – as any plausible moral theory must acknowledge -- we are agents who many times come thereby to have associated moral obligations to each other, and associated moral entitlements to demand. We thereby quite commonly come to have mutual moral obligations with respect to the shared activity; and we may well appeal to those obligations, when such there be, in our relevant practical reasoning. According to the basic thesis, however,

[209] Contrast with Margaret Gilbert, A Theory of Political Obligation at 153-6.
[210] This is not to deny that there might be, in certain cases, important moral considerations that are triggered by our shared agency.
[211] For appeals to the reinforcement of reliance see Facundo Alonso, "Shared Intention, Reliance, and Interpersonal Obligations," Ethics 119 (2009): 444-475.

appeal to such obligations is not essential to the explanatory structures, associated rationality norms, and practical reasoning at the heart of shared intention.

There is also a methodological point here. We do well to have a theoretical framework within which we can coherently pose the substantive question whether modest sociality quite generally involves such obligations. For example, we may want to ask whether the great apes, or young children, are capable of modest sociality without presupposing that an affirmative answer entails that the great apes, or young children, stand in relations to each other of mutual obligation. And that is what the basic thesis allows us to do.

2. Shared intention, Moral obligation

What then is the relation between shared intention and moral obligation? This is a question about moral obligations themselves, not about the participants' beliefs about their moral obligations. Further, there can be mutual moral obligations even if one or both of the participants intends not to comply with those obligations. You and I might each insincerely promise the other that he will help plow the commons. Since we have each promised – albeit, insincerely -- we each have a moral obligation to the other to plow. (Here I bracket issues about special circumstances that can block such obligations.) Nevertheless, each of us intends not to plow, even though, we can suppose, we each know of our obligation to plow. In such a case there is known mutual obligation without shared intention. We do not have a shared intention to plow together because we are not in a state that is set to explain our plowing. Rather, we are each in a state – intending not to plow -- that is set to explain why we do not plow together. Since, despite our (insincere) promises, we are not in a state that is set to explain our plowing together, ours is not a case of shared intention. So known mutual obligation does not ensure shared intention.

Suppose however that there is in fact a shared intention to plow together. This shared intention plays a characteristic explanatory role. And the claim of the basic thesis is that mutual obligations between the participants are not essential to that explanatory role, and so are not essential to shared intention.

Will there nevertheless generally be an associated mutual obligation to participate in the joint plowing (though a mutual obligation that is not essential to the shared intention)? Well, the basic thesis can and should acknowledge that the etiology of a shared intention will frequently include mutual assurances and the like, and that these will normally induce associated moral obligations. Shared intentions, once constituted, will normally issue in assurances and the like, and these will normally ground associated moral obligations. And once such moral obligations are on board, they may support the normal functioning of a shared intention – for example, by supporting its stability.

In saying this I am appealing to plausible views about when we come to have certain moral obligations. We need not sort out here the precise contours of the relevant principles of moral obligation. We can simply note that the complex elements and potential concomitants of a shared intention will many times engage relevant principles of moral obligation: they will may times satisfy the antecedent of a morally substantive principle the consequent of which says there are certain moral obligations.[212]

Indeed, it is open to the basic thesis to say that the elements of shared intention will in fact always engage some such substantive principles of moral obligation. What the basic thesis must deny is not that shared intentions are always accompanied by relevant moral obligations, or that these obligations can supplement relevant forms of practical reasoning and explanation of action. What the basic thesis must deny is, rather, that these obligations are essential to the explanatory role of shared intention.

That said, I think there are reasons to doubt that shared intentions are indeed always accompanied by such moral obligations. First, in some cases it seems strained to suppose that there is anything like an assurance. There are, for example, cases of brief and "accidental" sharing – as when you and I happen, unexpectedly, to walk

[212] In my "Shared Intention and Mutual Obligation" I try to specify at least some of the relevant moral principles by drawing on work of T. M. Scanlon. And this is, broadly speaking the tack that Facundo Alonso takes in his "Shared Intention, Reliance, and Interpersonal Obligations". While I continue to see this as a promising strategy, the view of modest sociality that I am developing here is open to a range of different approaches to such moral obligations. That is why I here refrain from tying my presentation to the specific features of a Scanlon-type account of these moral obligations.

together for a very short time and then part ways.[213] Again, in a case in which we more or less spontaneously applaud in response to a wonderful concert it seems plausible to say that our applauding is a shared intentional activity even though there are, at least initially, no relevant assurances, one to the other.[214] There is relevant interdependence in persistence, but not a background of mutual assurance.

Should we say that in such cases, though there is no assurance, there is a kind of induced or reinforced reliance that can ground relevant obligations?[215] Well, it seems that such induced or reinforced reliance is the source of an associated obligation only if the person who relies would suffer what T. M. Scanlon calls a "significant loss" if one did not follow through.[216] But in cases such as the ones just described there may well be no such threatened significant loss: what is at stake may not be important to any of the parties. But lack of importance need not block shared intention.

Of course, if one of the participants in one of these cases were to opt out without warning, the others may well be justifiably surprised since the prior evidence may well have supported an expectation to the contrary. But it is one thing to falsify a prior, epistemically justified expectation, another to violate a prior entitlement.

These remarks engage substantive issues in moral philosophy. But that is the point. Whether or not there is a certain moral obligation in a given case of shared intention is a complex moral issue, one to be settled by substantive moral argument.

Let me turn now to a second concern about a purportedly general connection between shared intention and such moral obligations. The concern is that some cases of shared intention will involve the kinds of special conditions that are plausibly seen as blocking such obligations. Let me quote here from my earlier discussion of this matter:

---

[213] This example comes from my "Dynamics of Sociality," Midwest Studies in Philosophy 30 (2006): 1-15, at 7. Facundo Alonso to some extent challenges my thought that there need be no mutual obligation in such a case; but his challenge is limited to, as he says, "cases of shared intention that, instrumentally or intrinsically, matter to us". However, as I go on to note in invoking Scanlon's concept of "significant loss," so long as there are cases of shared intention that need not "matter to us" in this way, Alonso's point is compatible with the main thrust of my use of this example. See Facundo Alonso, "Shared Intention, Reliance, and Interpersonal Obligations," p. 471 note 76.

[214] See above Chapter 00. I discuss this example further in "Dynamics of Sociality," at 7-8.

[215] See T.M. Scanlon, What We owe to Each Other (Harvard University Press, 1998), chap. 7; Facundo Alonso, "Shared Intention, Reliance, and Interpersonal Obligations," Ethics 119 (2009): 444-475.

[216] T.M. Scanlon, What We owe to Each Other (Harvard University Press, 1998), chap. 7. Facundo Alonso, "Shared Intention, Reliance, and Interpersonal Obligations," at p. 470.

> Perhaps two fans of Ayn Rand sing together even though each has made it clear to the other that neither is obligated to the other to continue: "no obligations," they each say. (Daniel Markovitz has observed that such cases are analogues of the doctrine of "employment at will" in American law.) Perhaps one of the participants coerces the other into participating in a shared activity. Perhaps we are engaged in a morally outlandish shared activity – ethnic cleansing, for example. It seems plausible to me that there are versions of each of these cases in which there is modest sociality without mutual obligation of performance.[217]

I continue to believe that in some such cases the most plausible thing to say is that there is shared intention without mutual obligation of performance, at least so long as the obligations in question are familiar sorts of moral obligations. But suppose I am wrong and one could persuasively argue that the structure of shared intention really does always ensure associated moral obligations. It does not follow that those obligations are essential to the basic explanatory roles of shared intention. We could still grant that the explanatory structures highlighted by the basic thesis are central to the explanation of coordinated action and planning in pursuit of a common end, and associated bargaining and shared deliberation. We could simply go on to grant that when such obligations are present they normally supplement these explanatory structures (if the participants care about conforming to their obligations).

3. Alonso on Shared Intention, Reliance, and Obligation

It will be helpful at this point to consider a subtle and instructive discussion of these matters provided by Facundo Alonso.[218] Alonso argues that a shared intention to J necessarily involves (a) the intentions of each of the participants in favor of J; and it also necessarily involves the further condition that (b) each participant relies on the other's both intending J and doing his or her part in J. Alonso then argues that in the

[217] "Modest Sociality and the Distinctiveness of Intention," at 151-2. Gilbert has responded to some such examples in "Shared Intention and Personal Intentions," Philosophical Studies 144 (2009): 167-87, at 177-179. My view is that examples such as these should give us pause about an overly tight connection to mutual obligation. But in this present discussion I am trying to highlight more systematic concerns.
[218] Facundo Alonso, "Shared Intention, Reliance, and Interpersonal Obligations," Ethics 119 (2009): 444-475.

circumstances of a shared intention each party will either intentionally or negligently induce or reinforce the other's cited (and necessary) reliance. And, drawing on work of T.M. Scanlon, he argues that this fact will, barring certain special circumstances (for example, forms of coercion), ground associated moral obligations: "failing special circumstances, shared intention generates relevant moral obligations between the participants."[219]

Let me offer six observations about how Alonso's proposal is related to the basic thesis. First, the basic thesis does not officially claim to have identified strictly necessary conditions for a shared intention. But it is compatible with the basic thesis that Alonso's condition (a) is, as he claims, necessary. Second, since the basic thesis offers sufficient conditions for shared intention, and these conditions make no explicit reference to reliance, it may seem that Alonso's appeal to (b) as a purportedly necessary element of shared intention is in tension with the basic thesis. But this is delicate. It might be that when the elements cited by the basic thesis function properly there will indeed be the kind of reliance Alonso cites. I will leave this as an open question.

Third, the obligations at the heart of Alonso's view are moral obligations. So here Alonso and I are in agreement (and we are both in disagreement with Margaret Gilbert[220]) that insofar as there is a close connection between shared intention and obligation, the relevant obligations are moral obligations. So it is no part of Alonso's view that the theory of modest sociality must introduce normative inter-relations that go beyond those grounded in the rationality norms highlighted by the basic thesis, taken together with relevant moral norms.

Fourth, Alonso argues that the cited conditions of shared intention ground such moral obligations, except in special circumstances. Recall, however, that my grounds for doubting that shared intention always brings with it such obligations included appeal

[219] 474

[220] Gilbert thinks there is a distinct kind of obligation, one she calls "obligations of joint commitment", and that these are not moral obligations. She writes: "moral requirements and the directed obligations of joint commitment differ radically in character." ("Shared Intention and Personal Intentions," Philosophical Studies 144 (2009): 167-187, at 184.) And see Margaret Gilbert, A Theory of Political Obligation (Oxford University Press, 2006) 156-63.

to cases involving such special circumstances (including cases of coercion, cases of immoral ends, and cases in which there is nothing important at stake and so no threat of "significant loss"). So there need be no deep disagreement between us on this point, though there may be disagreement about some particular examples.

Fifth, what is central to my view is the claim that we do not need to advert to such obligations to articulate the basic explanatory and rational structure that underlies the standard functioning of shared intention. This is the explanatory and rational structure that involves, roughly, interconnected and interdependent plan states in favor of the joint action by way of mutual responsiveness and meshing sub-plans, in a context of common knowledge. This claim of mine is compatible with the thought that if relevant moral obligations are in fact induced by the functioning of these plan states then these obligations will tend to support this social functioning (as least for agents who care about such obligations). But my claim is not compatible with the view that the explanation of the central forms of coordinated action and planning in pursuit of a common end, and associated bargaining and shared deliberation, must appeal to such mutual obligations. And it is important to note that such a claim about the necessity of such obligations for such explanations does not follow from Alonso's claims about the relation of shared intention to such obligations. Indeed, Alonso explicitly indicates that on his view "obligation-based guarantees are not necessary for shared intention to play its characteristic roles."[221] So there is agreement on this fundamental point.

Sixth, on the view I am defending the moral obligations that are associated with a shared intention are not themselves a constituent of that shared intention, but are, rather, concomitants of that shared intention. And here I take it Alonso and I are again in agreement. Alonso thinks that shared intention normally "generates relevant moral obligations". But this is not the idea that those obligations are a constituent of those shared intentions: "generates" is a relation between distinct elements.

4. Gilbert on Joint Commitment

[221] "Shared Intention, Reliance, and Interpersonal Obligations," at 463 n. 54. Alonso thinks that the needed stability of shared intention can be ensured by a "process of mutual reinforcement of the participant's intentions and relations of mutual reliance." See above note 000.

Our question has been whether the relation between modest sociality and mutual obligation poses a significant challenge to the basic thesis. I have been arguing that it does not. But now it is time to focus more carefully on work of a philosopher who explicitly disagrees, namely: Margaret Gilbert.

As Gilbert sees it, the move from individual to shared agency involves a move to a "joint commitment".[222] Gilbert does not try to provide an analysis of this idea of a joint commitment. She sees it as a basic, non-reducible idea, one that is the analogue in the case of shared agency to the idea of an individual, personal commitment -- as when an individual reaches an individual decision – in individual agency. Whereas the basic thesis tries to provide a broadly reductive theory of that in which "jointness" or "shared-ness" consists, the idea that a Gilbertian joint commitment is joint is, in effect, a primitive, non-reducible idea for Gilbert. Granted, Gilbert does in a way try to explain this idea further. For example, she says that "a typical context for the formation of a joint commitment of two people involves the parties in face-to-face contact mutually expressing their readiness to be jointly committed, in conditions of common knowledge."[223] But it is clear that this is no analysis, since the very idea of being jointly committed is appealed to in this remark.

It is worth pausing to appreciate this point. The approach I am developing aims to provide a substantive account of that in which the shared-ness of shared intention and shared intentional action consists. This shared-ness of shared intention consists, roughly, in the interlocking and interdependence of planning attitudes of each, planning attitudes whose contents favor the joint activity and the meshing roles of both, all in a context of common knowledge. The shared-ness of shared intentional action also consists in part in relevant mutual responsiveness of sub-intention and in action. In contrast, if we were to ask Gilbert what makes a commitment a joint commitment – in what does the joint-ness of the commitment consist – the answer would be that this joint-ness is a primitive, non-reducible phenomenon.[224]

---

[222] See e.g., her "What Is It for Us to Intend?" in her Sociality and Responsibility (2000) esp. pp. 21-24.
[223] "What Is It for Us to Intend?" at p. 21.
[224] In this respect, and as I have noted before, there is a kind of agreement between Gilbert and Searle: both appeal to a fundamental, primitive idea/phenomenon. However, whereas for Gilbert this primitive is a special relation between individuals, for Searle it is a special attitude of individuals. I seek to avoid either form of appeal to a practical primitive.

However, even given the non-reducibility of joint commitment, we can try to make certain substantive claims about such commitments. And that is what Gilbert aims to do. Gilbert proposes, first, that when – and only when -- there is such a joint commitment there is a "plural subject".[225] I will return in Chapter 7 to this idea, and to the question whether it adds anything of substance to the theory of joint commitment.[226] For now what is central is a second claim Gilbert makes, namely that "obligations with corresponding entitlements inhere in any joint commitment."[227] According to Gilbert, then, the step from individual to shared intention involves the introduction of this fundamental new phenomenon of "joint commitment", and thereby of a basic inter-personal normative relation that "inhere[s]" in it. And I take it that to say that these obligations "inhere" in a joint commitment is to claim that these obligations are an essential constituent of such joint commitments.

The claim is not that in modest sociality the participants believe that they have such obligations and entitlements. The claim is rather that in modest sociality there are these obligations and entitlements. Gilbertian joint commitments are a matter of a basic inter-relation between the participants; they are not just a matter of the beliefs of the participants about those relations. And there is an aspect of this idea with which the basic thesis agrees. Like Gilbert, the basic thesis sees modest sociality as essentiality involving certain inter-personal inter-relations: both theories reject the idea -- an idea that, as I noted earlier, seems to be endorsed by John Searle -- that what is essential to modest sociality is, at bottom, solely certain special attitudes of each of the participants. The difference between Gilbert's view and the basic thesis lies in the understanding of these inter-relations. On Gilbert's view, but not according to the basic thesis, what is fundamental includes mutual obligations and entitlements that "inhere" in the (non-reducible) joint commitment.

Gilbert's claim is not simply that, as a matter of normative fact, when there is shared intention there are associated obligations and entitlements across the participants. We can agree – and it is important -- that there are moral norms of civility

[225] "What Is It for Us to Intend?" at 22.
[226] As will emerge there, on the reading of this idea of a plural subject that is favored by Gilbert herself, it turns out that this claim about the connection between a joint commitment and a plural subject does not, in the end, add anything of substance to the account of joint commitment
[227] "What Is It for Us to Intend?" at 25 (italics in original).

and considerateness, norms that say that if you are engaged in a shared activity and you are going to opt out then, at least normally, you should at least alert the other participants and, perhaps, these participants are entitled to this from you (though there may be different conventions about such matters in different societies). There are moral norms of assurance-based and promise-based obligations that say, roughly, that if in a shared activity you have assured or promised that you will perform then the other participants are entitled to your performance. And there are also moral norms that say that if in a shared activity you knowingly lead the other participants to rely on your performance and thereby to be at risk of a significant loss if you do not perform, then you have some sort of associated obligation either to perform or to compensate or the like. As I have said before, the exact details of these moral norms are a matter of normative debate and reflection. But whatever we say in the end, to endorse such norms is not to agree with Gilbert that certain obligations and entitlements "inhere" in joint commitment, and so in shared intention. To endorse such moral norms is not to say in what shared intention consists; it is, rather, to be in a position to say that shared intention and shared activity frequently – or perhaps even always[228] -- engage some such norms and thereby induce associated obligations. I think it is clear that shared intention, shared activity, and its normal concomitants, do indeed normally engage one or more of these or other closely related moral norms. Indeed, the basic thesis is compatible with – though it does not require – the claim that there is always some such mutual moral obligation. What is crucial is that these are substantive claims about the interaction between shared intention and morality, not claims about that in which a shared intention consists.

What is in tension with the basic thesis is, then, not the idea that broadly moral norms of obligation and the like are, as Jules Coleman once put it (in conversation), "in the neighborhood". What is in tension with the basic thesis is the idea that certain mutual obligations and entitlements are necessary constitutive elements of shared intention, elements that always "inhere" in joint commitment, and so in shared intention.

Now, from the point of view of the basic thesis, the thought that mutual obligations always inhere in shared intention will seem an overstatement. First, and as I

[228] But see my discussion above in section 2.

have suggested by appeal to various examples in section 2, it is at least normatively unclear whether whenever there is shared intention there is mutual obligation. Second, when a shared intention – or its etiology, or downstream impact -- does ground relevant mutual obligations it will seem, from the point of view of the basic thesis, that the shared intention and the mutual obligations are distinct elements, the former – and perhaps its concomitants -- grounding the latter.[229] And so it will seem incorrect to say that the obligations "inhere" in the shared intention. And third, on the assumption (one it is the burden of this work to defend) that the basic thesis succeeds in providing a sufficiently rich model of the forms of interpersonal interconnection and interaction characteristic of modest sociality, we do not need to appeal to such mutual obligations in order to provide an explanatorily adequate account of those interconnections and interactions.

Let me try to advance the discussion by considering Gilbert's recent essay, "Shared Intention and Personal Intentions,"[230] where she offers the following claims:

a. "Intuitively an appropriate agreement between the parties is sufficient to bring a shared intention into being."[231]
b. "an adequate account of shared intention is such that it is not necessarily the case that for every shared intention, on that account, there be correlative personal intentions of the individual parties."[232] [Gilbert sees the intentions of the participants that are cited in the planning model of shared intention as "correlative personal intentions".] Gilbert calls this "the disjunction condition".
c. The disjunction condition is defended by way of an example in which Olive reports "Our plan was to hike to the top of the hill. We … started up. As he told me later, Ned realized early on that it would be too much for him to go all the way to the top, and decided that he would only go half way. Though he no longer had any intention of hiking to the top … he had as yet said nothing about this to me …[Before halfway] we encountered Pam who asked me how far we intended to go. I said that our intention was to hike to the top ..,

[229] Or, as Alonso would say, generating the latter.
[230] "Shared Intention and Personal Intentions," Philosophical Studies 144 (2009): 167-187
[231] 169. This claim is restated at p. 172 where, in note 17, Gilbert makes it clear that talk of "appropriate" is aimed at insuring that the agreement concerns what "to do".
[232] 172

as indeed it was."[233] Gilbert thinks Olive speaks truly in saying "as indeed it was", and that this supports the disjunction condition.

d. "an adequate account of shared intention will entail that, absent special background understandings, the concurrence of all parties is required in order that a given shared intention be changed or rescinded…"[234] Gilbert calls this "the concurrence condition" (and she also supports this with an example).

In reply, I do not accept any of these claims (as naturally interpreted). In particular, it seems to me that the plausibility of the purported disjunction and concurrence conditions on shared intention comes from not sufficiently distinguishing shared intention and mutual obligation.

Begin with claim a., the claim that "an appropriate agreement between the parties is sufficient to bring a shared intention into being." Perhaps this claim has some intuitive support. But if we are really talking about shared intention of a sort that is set to explain joint activity, the claim seems to me false as a general claim. This is because – as I have emphasized -- people can insincerely agree to, say, plow the fields together, even though each participant fully intends not to act in accord with that agreement. Such insincere agreements seem all too possible. In such cases of insincere agreement there normally is, to be sure, a normative structure of obligations and entitlements. So we may agree that "intuitively an appropriate agreement between the parties is sufficient to bring a [*mutual obligation*] into being." But, given the possibility of insincerity, an agreement seems not to ensure that there is a shared intention of a sort that explains joint action. And a problem with an appeal to intuitions here is that intuitions can mislead, by not keeping track of the distinction between shared intention and mutual obligation.[235]

Could Gilbert insist that agreements – even when insincere – do ensure shared intention since, after all, they do ensure mutual obligation? Well, you can use the words

[233] 171-2
[234] 173
[235] Could one respond by saying that there is, strictly speaking, no agreement if there is the cited insincerity on the part of either participant? Well, if this is how we understand the idea of an agreement then we cannot appeal (as Gilbert does at p. 172) to the phenomenon of agreement to support the disjunction condition, since agreements in this special sense do not conform to the disjunction condition. After all, agreements in this special sense would require that each participant intend to comply.

"shared intention" here if you want. But then we should be clear that shared intentions, so understood, do not ensure the intention-based motivational basis of modest sociality, since they do not ensure relevant intentions to act. (Though the recognition of such an obligation can motivate, it presumably does not adequately motivate such insincere parties to the agreement.) But in the absence of this intention-based motivational basis we do not yet have a phenomenon that can play the basic explanatory role that, as I have been supposing, shared intention is to play. So, at the least, this would not be shared intention in a sense that would do the explanatory work of shared intention, as I have been understanding that work.

Turn now to b. and the disjunction condition. It is clear that the planning model of shared intention, as articulated in the basic thesis, would not agree with the disjunction condition, since that planning model sees shared intention as consisting of what Gilbert calls "personal intentions". (Though these are "personal intentions" with social contents.) So what is the argument for the disjunction condition? Well, the basic argument is the example in c. But it seems to me that insofar as we are willing to agree with Olive, what we are thinking is that even after Ned changes his mind Olive and Ned have a mutual obligation to climb the hill to the top. As I see it, once Ned has changed his mind they do not any longer have the kind of shared intention to climb to the top that is set to explain their climbing to the top. After all, at that point Ned fully intends not to climb to the top, and what is now going to need to be explained is not their climbing to the top, but rather their failure to climb to the top. Granted, Olive's cited belief, after Ned changed his mind but before Ned had told her of this change of mind – her belief that they have a shared intention to climb to the top -- might have been epistemically justified. But we can grant that this belief is epistemically justified without granting that it is true.

A similar point can be made concerning claim d., the claim that "the concurrence of all parties is required in order that a given shared intention be changed or rescinded". It seems to me that at most what is true here is that concurrence is needed in order to cancel the mutual obligations. Though Ned cannot unilaterally cancel relevant obligations, it is nevertheless true that once Ned fully intends not to

climb to the top he and Olive no longer have a shared intention to climb to the top, in a sense of shared intention that gets at the basic explanatory phenomenon.

So I think that while the concurrence condition and the disjunction condition are plausible conditions on mutual obligation, they are not plausible conditions on shared intention, understood as a basic explanatory factor.

5. Normativity, sociality, and Ockham's Razor

I have said that the basic thesis does not make an essential appeal, at the ground level, to the purported, relational normative phenomena highlighted by Gilbert's account of joint commitment. But we need to understand this contrast with care. It would not be accurate simply to say that the basic thesis eschews appeal to the normative. After all, it is central to the planning theory that there are norms of individual intention rationality. It is, further, a central claim of the basic thesis that within relevant structures of interconnected planning agency these norms induce associated norms of social rationality. And the basic thesis can emphasize that shared intention, as it understands it, will interact with morality. Instead, the relevant issue between the basic thesis and Gilbert's theory concerns whether our theory of shared intention must advert, at the ground level, not only to the cited norms of intention rationality but also to the distinctive kinds of mutual obligations cited by Gilbert.

Why might one think that such mutual obligations must come in at the ground level? Well, return to cases of mere concatenation of activities with mutual tracking and adjustment. We can suppose that such cases take place within a context of common knowledge. It is, for example, common knowledge, between me and the stranger, that we are walking near to each other and in the same direction and at roughly the same pace down the street. And that is why we keep an eye out to avoid collisions. But ours is not a case of shared intentional activity. Why not? It is not a matter of the absence of knowledge of each about each. Here Gilbert and I are in agreement. Instead, there is in the case of shared intentional activity a distinctive practical tie, one that is not ensured by further merely epistemic conditions. But what could this practical tie be? And here, as I have noted before, it is tempting to say: these are, at least in part, ties of obligation of each to each.

This last step supposes that if the practical ties are not merely an epistemic matter then they are, at least in part, a matter of ties of mutual obligation. But what the basic thesis helps us to see is that our philosophical options are richer than this. In particular, once we have on board the planning theory of individual agency we have the resources – conceptual, metaphysical, and normative – to characterize, without appeal to mutual obligations (and so without appeal to Gilbertian joint commitment), both intentions with distinctive contents and distinctive practical ties of interlocking and interdependence between these intentions of the participants. Granted, when there are relevant mutual, moral obligations – as there commonly are -- their recognition will normally contribute to the standard functioning of shared intention, not least because of the normal impact of the recognition of such obligations on the stability of the shared intention. And there will be cases in which the prior concerns of the participants diverge in ways that make a stable shared intention unlikely without the support of such mutual obligations. In such cases we will frequently have good reason, in support of our joint activity, to try to arrive at obligation-creating mutual assurances, agreements or promises. Nevertheless, the conjecture of the basic thesis is that, given its resources, we can characterize the fundamental practical ties without essential appeal to such mutual obligations.

This is not to deny that there is an important social phenomenon in which different agents are tied together in their pursuit of a common end in part by relevant moral obligations. Common versions of such cases involve moral obligations grounded in promises, assurances, intentionally induced reliance, and the like. These are pervasive and fundamentally important forms of human sociality. Shared intention and modest sociality frequently activate relevant moral pressures, including especially such moral obligations. But there is no reason to say that these obligations “inhere” in shared intention.

What about Gilbert-type joint commitments? Should we say that in addition to the phenomenon of modest sociality that I have been characterizing, and in addition to garden-variety moral phenomena involving mutual moral obligations or other familiar moral inter-relations, there is yet a further and distinct phenomenon in which parties are related by Gilbert-type joint commitments? This is a different question than the question

of whether there is – as there clearly is -- a phenomenon of activity structured by mutual moral obligations. After all, joint commitment is supposed to be a distinct and unanalyzable phenomenon, not just a special kind of moral obligation. So we need to know if we really do need to appeal to this further practical primitive, over and above both morality and the kind of modest sociality I have characterized. And here Ockham's Razor counsels caution. We need to know why we need to appeal in our theory to this yet further practical primitive. And my argument has been that we do not find grounds for the introduction of this further practical primitive in the consideration of garden-variety modest sociality. We can understand such sociality along the lines of the basic thesis; and we can then go on to explore the different ways in which cases of modest sociality engage substantive moral norms, including norms of civility and of obligation.

This Ockham's-Razor argument generalizes. Once we have the resources of the planning theory we can construct a rich model of modest sociality along the lines of the basic thesis. This model provides resources to understand and explain complex forms of social functioning and social rationality involved in modest sociality. We can then go on to embed such modest sociality within contexts that engage a range of substantive moral norms. We thereby arrive at a model for an even wider range of forms of sociality. A theorist who nevertheless thinks – as do, in different ways, Gilbert and Searle – that we still need a new practical primitive, needs to argue that we have so far still failed to capture important forms of small-scale sociality. At this point, the burden of argument seems to me to be on such a theorist.[236]

[236] In his "Shared Agency and Contralateral Commitments," The Philosophical Review (2004): 359-410, Abraham Sesshu Roth seeks a theory that is broadly in the spirit of Gilbert's appeal to special mutual obligations but appeals only to a somewhat weaker normative inter-relation he calls "contralateral commitment". Roth thinks that "contralateral commitments" between the participants in shared agency fall in the space between the rational commitments that are explained by a theory along the lines of the basic thesis, and the kinds of inter-personal moral ties explained by appeal to substantive moral principles of obligation and the like. This claim has the right form to be a response to the generalized Ockham's-Razor argument I have just articulated; and it has an advantage over Gilbert's version in making a more modest normative claim. That said, I myself am skeptical that there is a robust inter-personal normative phenomenon that lies in the cited space. But I will not try to settle this matter here. (Roth himself goes on to claim that to explain this special contralateral commitment we need to appeal to the difficult idea that an agent can act "directly" on the intentions of another agent. If my skepticism about the cited substantive claim is correct, we need not be led in this way to appeal to that difficult idea.)

Chapter Seven: Group Agents Without Group Subjects

In modest sociality, according to the basic thesis, the participants are inter-connected planning agents. Given appropriate planning attitudes, and interconnections across those attitudes, it will be true that, say, we intend to paint the house together, and, if all goes well, that we do indeed paint the house together as a shared intentional activity. In expressing each of these claims, 'we' functions as a grammatical subject. Should we also think that in such cases 'we' refers literally to a group agent who paints? to a group subject who intends to paint?

1. Group agents and the basic thesis

Begin by returning to Petersson's essay. In Chapter 5 I discussed Petersson's argument that my account did not provide a strong enough model of modest sociality. I concluded that Petersson's objection, as well as a similar objection from Gold and Sugden, did not succeed, as we could see once we fully appreciated the distinctiveness of intention. Nevertheless, we can learn from reflecting on the amendment to my account that Petersson goes on to offer.

Petersson asserts that our theory of, as I call it, modest sociality needs a notion of collective activity that does not involve shared or collective intention but that does conceive of the group as itself a "causal agent".[237] To think of the group as a causal agent, in Petersson's sense, one need not see it as the subject of intentions or the like. Talk of "agent" here is not intended to go beyond the idea of an internally structured locus of causal powers and of the causal attributions of effects. As Petersson says, "this way of speaking simply places the object in a certain causal role, and refers to an effect for which internal features of the object is a condition."[238] To see a group as a causal agent, in the relevant sense, one need only see it as having sufficient internal organization such that, because of that internal organization, we can reasonably attribute to that group "causal powers or dispositions,"[239] and can reasonably attribute

[237] "Collectivity and Circularity," at 148. Petersson credits Tim Crane for this term. See Crane, "The Efficacy of Content," in Jan Bransen and Stefan E. Cuypers, eds., Human Action, Deliberation and Causation (Boston: Kluwer, 1998): 199-225, at 220.

[238] "Collectivity and Circularity," at 149. Petersson emphasizes the parallel with non-intentional agency of an individual: "If you point out that your jiggling made me spill coffee, my causal role for that effect is stressed." (p. 149. The first emphasis is in the text, the second has been added.)

[239] "Collectivity and Circularity," at 148.

certain upshots to it as something it, the group, causes.[240] For example, we may see a swarm of bees as such a group causal agent:

> We may … watch the swarm as one causal agent and think about what it might do, wonder what makes it fly this way rather than that, think that some of its acts seem unexpected, and so on. … What makes you regard the swarm as the unit of causal agency need not be any specific knowledge about its internal structure, but just that its behavior gives you reason to think there is some such structure.[241]

This idea in hand, Petersson goes on to propose that this "weak notion of collective activity [involving a weak idea of a group causal agent] must figure in the content of the intentions of the parties to a collective action."[242] And here he sees himself as going beyond the account of these contents that I have provided.

It is central to Petersson's proposal that these ideas of collective activity and group causal agency can appear in the contents of the intentions of each participant, in the account of shared intention, without falling prey to a circularity objection; for the relevant idea of a group causal agent does not involve the idea of shared intentionality. After all, swarms can be group causal agents in Petersson's sense, as can, I take it, flocks, or certain human mobs or crowds in the absence of a shared intention. In this respect, Petersson and I have similar theoretical ambitions: we each seek an account of the contents of the intentions of the individuals that help constitute the shared intention, an account that does not fall prey to an unacceptable circularity. However, Petersson's concept of collective activity with a group causal agent is somewhat stronger than the

---

[240] This is in the spirit of Donald Davidson's observations concerning the relation between "the accordion effect" (a term due to Joel Feinberg) and agency. Davidson says, for example, that "an agent causes what his actions cause"; and he indicates that this is a "mark of agency". (See Donald Davidson, "Agency" in his Essays on Actions and Events at 53-54.) Here Davidson's view seems to be that if X causes what its actions cause then X is an agent (and its actions are thereby accurately so-called). And Petersson's idea is that a group is a group agent when it causes what its actions cause (and its actions are thereby accurately so-called). The swarm of bees, for example, causes what its stings cause. (If the victim dies then the swarm causes the victim's death.) If we interpret Davidson along such lines, however, then there will be a tension with his view, to be discussed below, about the necessary relation between agency and intentional agency.

[241] At 152-3. I have omitted parts of this passage that are tied to the further point that we might also refrain from seeing the bees in this way, and even if we do see them this way we might have different substantive theories of the internal structure. While I would also agree with these further points, they can safely be omitted here.

[242] "Collectivity and Circularity," at 155. Emphasis added.

one I have used for this purpose at the most basic level of my construction.[243] This is because Petersson's concept explicitly brings with it the idea of the group as causal agent, whereas, roughly speaking, mine makes do with, as Petersson puts it, "the notion of a mere set of intertwined acts".[244](146) Petersson says that his claim is

> stronger than Bratman's initial proposal that it is sufficient that the act description, as it figures in the content of the agent's intention, merely satisfies the behavioral conditions for a joint activity. Such a description would be neutral with respect to causal agency … My additional requirement is that the notion of the team as the causal agent must enter that content.[245]

Now, I take it that in appealing to the idea of a group causal agent (in the absence of group intentions or the like) the thought is not that this form of causation is metaphysically distinct from the underlying causal processes involving the organized elements of the group. The thought is, rather, that it can be true in this weak sense that the group is a causal agent, but when this is true this truth consists in facts about the relevant underlying structure of inter-related individuals and casual processes that are shaped by that organizing structure. This is a metaphysically modest idea of a group causal agent, one that is compatible with an underlying metaphysics of inter-related individual agents. Nevertheless, we can see Petersson's proposal as an independent – albeit, limited -- challenge to the conceptual adequacy of my account: does my account of modest sociality need, at bottom, an idea of the group as a causal agent, an idea that goes beyond the conceptual resources of the planning theory of individual agency?

I think not. I agree with Petersson that when there is, as I have called it, modest sociality, we can plausibly see the participants as together constituting a group that, because of its internal organization, is a bearer of causal powers and something to

---

[243] Recall that in Chapter 2 section 1 I distinguished between a concept of our activity that is (a) neutral with respect to shared intentionality, and one that (b) can be articulated using the conceptual resources of the planning theory of individual agency. In this discussion of Petersson's proposal I am returning, as promised, to the issue of whether our theory can work, at the basic level, with a concept of our activity that satisfies both (a) and (b), or must instead retreat to a concept that satisfies (a) but not (b).

[244] "Collectivity and Circularity," at 146; though it is important that I would say: intertwined individually-intentional acts. See next footnote.

[245] "Collectivity and Circularity," at 155. Emphasis added. Petersson says here that the kind of action description I have in mind can include only "a behavioral description like 'their arms move'". But as I have noted earlier, this is not correct. I assume the descriptions will involve concepts of individual intentionality. What is crucial is that they not involve the concept of shared intentionality. In the discussion to follow I will assume that this correction has been made.

which certain effects can be causally attributed. So we can see them as constituting a group causal agent in Petersson's (weak) sense. That the house is now painted can, for example, be causally attributed to us. But this does not show that we need, in the foundations of the theory of modest sociality, a concept of a group causal agent that goes beyond the conceptual resources already provided by the planning theory of individual agency and that "must" appear in the relevant intention-contents.

Consider what Petersson says about what it is to think of the group as a causal agent:

> In regarding the group as the causal agent, we imply that there is some glue – there is something about the intrinsic features of the group and about the participants' role in the base of the group's causal powers, which distinguishes members from nonmembers – although we refrain from specifying this glue.[246]

Petersson supposes that if one does believe that there is some internal organization that provides the appropriate sort of "glue" then one regards the set of participants as constituting a causal agent.

Now, according to the basic thesis, what glues together the individual participants in central cases of modest sociality – in contrast with swarms or flocks or certain kinds of crowds -- are the cited social-psychological ties. It is, in a basic case, this social-psychological structure that explains the organization of their activity. Given this social-psychological structure and this organized activity we can then go on to talk of the group as a causal agent that is constituted by its participants, organized in these social-psychological ways. This talk of the group as causal agent is built on our appeal to this social-psychological structure. As I see it, however, we do not need this idea of a group causal agent to articulate this structure in the first place.

This is not to deny that there is a concept of a group causal agent that can be explained independently of explicit appeal to the particular social-psychological structures highlighted in the basic thesis. We can follow Petersson and introduce a higher-order notion that involves existential quantification over potential forms of relevant glue: to say that there is a group causal agent is to say that there is a form of internal organization that appropriately explains relevant causal upshots; and we can

[246] "Collectivity and Circularity," at 153.

say this without specifying exactly what that internal organization is and, in particular, without referring specifically to the social-psychological structures highlighted in the basic thesis. Examples like those of swarms of bees, or flocks, or certain kinds of crowds, suggest that such a concept would be theoretically useful. Since this concept involves existential quantification over forms of social organization, it does seem, strictly speaking, to go somewhat beyond the conceptual resources of the planning theory of individual agency (though, as emphasized, it remains neutral with respect to shared intentionality). What I deny, however, is that it is necessary to include this concept within the contents of relevant intentions of individuals in order to specify the basic social psychological ties at issue in modest sociality.

Of course, if the participants have the attitudes cited by the basic thesis, then it follows "that something glues [the relevant] components together".[247] This is simply existential generalization. On the assumption that Petersson has given us sufficient conditions for regarding a "set of objects as one causal agent,"[248] it follows that the participants who satisfy the conditions of the basic thesis are in a position to see their group as a causal agent in Petersson's (weak) sense. And they are each in a position to see that if their intentions are realized there will be a joint activity of which the group is a causal agent. But it does not follow that the contents of their intentions in the basic case must include a further, primitive idea of a group causal agent.

In contrast, it is Petersson's view that if these contents do not include this further idea of a group causal agent then the theory will not be able to distinguish (what I call) modest sociality from strategic interaction. This is the challenge that I considered in Chapter 5. But, as I argued in that chapter, once we recognize the distinctiveness of intention we can see that this challenge fails.

This is to reject Petersson's challenge to the conceptual underpinnings of the basic thesis, but to endorse the idea that the metaphysics of modest sociality may include group causal agents (in the cited, weak sense of causal agent). However, this appeal to group causal agents is to be understood in a sense that identifies these groups with a structured complex of the participants, and identifies their causal role, in

[247] 151
[248] 151

any particular case, with relevant causal roles of those participants suitably inter-related. And the basic thesis provides a model of the distinctive glue that is characteristic of those cases of group causal agency that are, in particular, cases of modest sociality.

According to the basic thesis, in modest sociality the contents of the intentions of each involve the concept of "we". In basic cases this is either simply the distributed "we", or a concept of "we" (such as "those in this room") that picks out a group of people in ways that do not involve appeal to the very idea of shared intentionality. However, if all the conditions of the basic thesis are satisfied it will follow that these several, distributed participants are interconnected – social-psychologically glued together -- in a way that makes it true that they together constitute a group causal agent, one to which various effects can be causally attributed. Further, they are tied together in ways characteristic of, in particular, shared intentionality. Though the concept of 'we' involved in the contents of the intentions of each need not, in the most basic cases, invoke the very idea of a group causal agent, the basic thesis describes a world in which there are (in the cited, weak sense) group causal agents involved in modest sociality. These group causal agents are not mere collections of the several, distributed participants, though they do consist in those participants organized and inter-linked in the ways highlighted by the basic thesis. In this sense, while the basic thesis is conceptually conservative about the concept of 'we' that needs to be employed in our basic construction of modest sociality, it is metaphysically accommodating concerning the group agents that are causally involved in modest sociality. It can say that the agent of, say, the duet singing is the group, and it is the structure of the participants that is characterized by the basic thesis that constitutes the group that is that agent. Since such group agents are identified with appropriately interconnected structures of individual agents – where these are plan-theoretic interconnections -- this is a form of metaphysical accommodation that is compatible with the continuity thesis.

Once we see this we can allow that the parties to a shared intention may themselves go on to conceptualize what they intend in a way that involves the idea of a group causal agent, understood in terms of existential quantification over appropriate forms of social organization. This kind of conceptual ratcheting is familiar from our earlier discussions of worries about circularity in the philosophy of action. But such

conceptual ratcheting is compatible with the conceptual conservatism of the basic thesis, since the concept of ”we” at work in the basic cases that ground this ratcheting need not itself involve the very idea of a group causal agent.

2. Group Subjects?[249]

The basic thesis sees shared intentions as consisting in a complex state of affairs, one that involves inter-connected attitudes of the participating planning agents. Now, when there is an intention of an individual agent there is an individual subject who so intends. So we need to ask: when there is a shared intention is there a group subject who so intends?

Consider Jones who, alluding to his partner Smith, says:

1. We are singing the duet together as a shared intentional activity, and

2. We intend to sing the duet together.

According to the basic thesis, 1. and 2. would be true if there is an appropriate social-psychological web, and this web connects up in the right way to action. Note however that, so far, the conditions that the basic thesis provides as sufficient for the truth of 1. and 2. do not explicitly say to what ‘we’, as it appears as the grammatical subject in 1. and 2., refers.

Granted, I have emphasized that according to the basic thesis the concept of ‘we’, as it is involved in the contents of the relevant intentions of Jones and Smith, can be a merely distributed notion – it can be merely the idea of a collection of the several individual agents. But I also noted in the previous section that if 1. and 2. are true in the way envisaged by the basic thesis then there will in fact be a group causal agent that is not merely a concatenation of the individual agents, but is a structured complex of those individuals.

The idea is that if 1. is true in the way envisaged by the basic thesis then there is in fact an interconnected collection of the individuals that, because of its internal organization, will count as a group causal agent of the action reported. So if 1. is true in the way envisaged by the basic thesis then there is this group causal agent and that group agent can in fact be the referent of ‘we’ in 1. This group – the organized structure

[249] Some of the thoughts in this section derive from my reactions to detailed and very helpful comments, both oral and written, from Carol Rovane.

involving Jones and Smith – is the agent of the shared action described in 1. And this is to say that 'we' in 1., when 1. is true in the way envisaged by the basic thesis, can refer not simply to the distributed collection of Jones and Smith but rather to the group (which consists of the internally organized structure involving Jones and Smith), and that this group is the agent of the shared action reported in 1.

Further, according to 1. this group is the agent of, in particular, a shared intentional activity: this is not like the agency of a swarm of bees or a panicky crowd. And what makes this a shared intentional action is in large part the shared intention reported in 2. Should we say then that 'we' in 2. is not only the grammatical subject, but also refers to the group agent referred to in 1. and says that this group is the subject of this intention?

Here I think that so long as we are focusing on cases of modest sociality our answer should be that this is not in general true: in modest sociality there need not be a group subject who has the shared intention. To talk of a subject who intends is to see that subject as a center of a more or less coherent mental web of, at the least, intentions and cognitions. The idea of a subject who intends X but has few other intentional attitudes – who intends X in the absence of a mental web of that subject in which this intention is located -- seems a mistake. This is a lesson we can learn from Donald Davidson's work on the holism of the mental.[250] But, as noted in Chapter 4, in cases of modest sociality the sharing will typically be very partial and very limited: Jones and Smith might have no other shared projects before them, and might significantly diverge in the reasons for which they participate in this shared project and in their relevant judgments of the right and the good. The sharing can be quite transitory: this might be a very short duet. And the sharing can cross-cut: Jones might sing this duet with Smith while playing chess with Brown. These features of the sharing need not block the idea of an internally organized group causal agent of a shared intentional activity, an agent that is limited in its causal impacts and, perhaps, quite temporary. But the moderate holism of subject-hood distinguishes this idea of a causal agent of a shared intentional activity from the idea of a subject of a shared intention.

[250] Donald Davidson, "Mental Events," in his Essays on Actions and Events second edition (Oxford University Press, 2001): 207-225. And see also Carol Rovane, The Bounds of Agency: An Essay in Revisionary Metaphysics (Princeton University Press, 1998).

In modest sociality it seems plausible that the group – a structured collection of individuals -- is, in a weak sense, the agent of the shared intentional activity. And shared intentional activity involves shared intention. But when we turn to the idea of the subject of the shared intention we should not expect that in modest sociality there will in general be a sufficiently robust, coherent web of relevant shared attitudes to support the claim that the group is that subject. In this way, being the agent of the shared action can come apart from being the subject of the shared intention, even given that the shared action is organized by shared intention.

Davidson's idea of the holism of the mental is, roughly, that we make sense of ascriptions of contents and attitudes to the same person in ways that require that these contents and attitudes more or less hang together in characteristic ways. We also owe to Davidson the idea of a tight connection between being an agent and being a subject of relevant mental attitudes. In his essay "Agency" Davidson ties agency to intentionality: "a man is the agent of an act if what he does can be described under an aspect that makes it intentional."[251] And Davidson's theory of the intentionality of action sees such intentionality as, roughly, explainability in the relevant way by the agent's relevant attitudes. But there are these relevant attitudes of the agent, according to Davidson, only if these attitudes are embedded in a holistic, and more or less coherent mental web of attitudes. It follows that, according to Davidson, to be an individual agent involves being an individual intentional agent, and so an individual subject of a holistic web of attitudes.[252]

When we turn to shared intentional agency, however, this tight connection between agency and subject-hood does not survive. This is not to deny that there may be special cases in which it is plausible to talk of a Davidsonian group subject.[253] The

[251] Donald Davidson, "Agency" in his Essays on Actions and Events. Esp. p. 46.
[252] As indicated above in note 00, there is a tension between this Davidsonian theme and Davidson's idea, in this same essay, that "we may take the accordion effect as a mark of agency"(54) since the accordion effect is tied to matters of causation, rather than to matters of intention. Davidson seems aware of this tension when he insists that "the accordion effect is not applicable if there is no intention present." (56) As I see it, however, his brief argument for this claim is not persuasive. Indeed, Petersson's example of his spilling of the coffee (above, note 00) suggests a counter-example: If his spilling the coffee ruined his friend's suit then he, Petersson, ruined the suit; so it seems we have the accordion effect even if "there is no intention present."
[253] This possibility is the central concern of Christian List and Philip Pettit, Group Agency: The Possibility, Design and Status of Corporate Agents (2011). And for a very suggestive, earlier treatment of related

claim is only that there can be shared intentional agency, and associated shared intention, in the absence of a social subject of that shared intention, and that this is indeed the normal case of small-scale modest sociality.

I think, then, that in basic cases of modest sociality the referent of "we" in 2. should be understood by the basic thesis as different from the referent of "we" in 1. If 1. is true in the way envisaged by the basic thesis then we can see 'we' in 1. as in fact referring to the internally organized group causal agent of the shared intentional activity, an agent whose structure is articulated by that thesis. That agent is a causal source of the shared intentional activity (which is not to say that the participants must include this idea of a group agent in the content of their relevant intentions). In contrast, if 2. is true in the way envisaged by the basic thesis then there is an appropriate interpersonal social-psychological web that makes 2. true. But even given that social-psychological web, it may well be that there is no group subject of the shared intention, given a moderately holistic notion of such a subject. So it will normally be a mistake to think of 'we' in 2. as referring to such a subject. Rather, we can see 'we' in 1. as referring to a group causal agent (in the cited, weak sense), but not see 2. as saying that this group is the subject of the shared intention.[254] If 2. is true then it is true that Jones and Smith together intend to sing; but there may be no group that is the (holistic) subject of that shared intention to sing.

---

ideas see Carol Rovane, The Bounds of Agency: An Essay in Revisionary Metaphysics. List and Pettit argue that in certain cases issues raised by "discursive dilemmas" will sensibly lead members of a group to seek to construct overall group positions in a way that bears only a complex relation to the attitudes of each of the individuals and that conforms to conditions of a Davidsonian group subject. However, List and Pettit also explicitly allow for the kind of modest sociality that is the target of my theory here – indeed, they build their story of group agency and group subject-hood on the top of structures of (what I am calling) modest sociality, structures that themselves need not involve a group subject. They call such modest sociality cases of joint intention and joint action. And they write: "a group of individuals may … form a joint intention to become a group agent. They each intend that they together act so as to form and enact a single system of belief and desire, at least within a clearly defined scope." (34) And I take it that on their view, such an originating joint intention need not itself have a group subject, since they explain group agents and group subjects as the outcome of such joint intentions. Since my target is limited to such modest sociality I will not try here to assess their claim that in some special cases, ones in which discursive dilemmas loom large and there are practical pressures to resolve them in a systematic way, there really does emerge a Davidsonian group subject.

[254] Kirk Ludwig writes that "plural agents are puzzling. They must, it seems, have beliefs, desires and intentions." On my view, in contrast, there can be group agents who are not the subject of the relevant shared intentions. See Kirk Ludwig, "Collective Intentional Behavior from the Standpoint of Semantics," NOUS 41 (2007):355-93, at 357.

This returns us to a point from Chapter 4 about a Lockean aspect of both individual and shared agency. The basic thesis highlights both referential interlocking and commonality of content of the intentions of the different participants in modest sociality. In this respect, according to the basic thesis, the social glue characteristic of modest sociality at a time to some extent parallels the Lockean structure of individual planning agency over time. This social glue is quasi-Lockean. And this might make the idea of a group subject in modest sociality more tempting. But I think that this would be a mistake. Even given these inter-personal quasi-Lockean inter-relations in modest sociality, it still seems incorrect to suppose that there is, in general, a (holistic) group subject in such modest sociality.

Granted, if a small group has a shared intention to bring about an untoward effect and succeeds in doing that, we may want in some sense to hold that group accountable. But then what we need, if we want to defend this, is an interpretation of such accountability (something I will not attempt here) that does not require a group subject.

I noted earlier that in Margaret Gilbert's view whenever there is shared intentional activity there is a "plural subject" of the involved shared intentions.[255] I have been arguing that in standard cases of modest sociality what makes it true that there is a shared intention does not suffice for there being a subject of the shared intention, given the holism of subjecthood. Am I thereby in disagreement with Gilbert?

It depends on how we are to interpret Gilbert's talk of a plural subject.[256] On a robust interpretation, Gilbert's talk of a plural subject is sufficiently analogous to our talk of an individual subject that it engages the idea of a center of a moderately holistic mental web.[257] And my claim is that there need not be a plural subject, in this robust sense, for there to be modest sociality. So if this is how we are to interpret Gilbert, then Gilbert and I are indeed disagreeing. On a more deflationary reading, however, Gilbert's talk of a plural subject is only a facon de parler – a shorthand for talk of a set of

[255] "What Is It for Us to Intend?" at 19, 22.
[256] J. David Velleman notes this interpretive issue in his "How to Share an Intention," as reprinted in his The Possibility of Practical Reason at 201. Velleman himself seeks a theory in the spirit of the first, more ambitious interpretation.
[257] Philip Pettit and David Schweikard interpret Gilbert in this way in their "Joint Actions and Group Agents," Philosophy of the Social Sciences 36 (2006): 18-39, at 32.

persons who are, in her sense, jointly committed with respect to a specific joint action. The substantive metaphysics of modest sociality lies entirely in such joint commitments – commitments that can be quite local and quite limited. So interpreted, Gilbert is not claiming that there is a plural subject over and above specific joint commitments, in the way in which there is an individual subject over and above specific intentions of that subject. On this interpretation, the appeal to the idea of a plural subject does no further philosophical work in Gilbert's theory, and Gilbert and I are not disagreeing about the need for a plural subject in shared intentionality. Our disagreement is, rather, about how precisely to understand the specific interrelations among participants that constitute specific cases of shared intentionality.

Now, in correspondence Gilbert has indicated her preference for this second reading, citing her A Theory of Political Obligation, where she says:

> It is useful to have a label for those who are jointly committed with one another in some way. I have elsewhere used the label 'plural subject' for the purpose and shall use it that way here. To put it somewhat formally: A and B (and...) (or those with feature F) constitute a plural subject (by definition) if and only if they are jointly committed to doing something as a body---in a broad sense of 'do'."[258]

So that is how I will understand Gilbert's view. But we should note that, if this is indeed her view, then we need to go back to a point, in Chapter 6, about our understanding of the very idea of joint commitment. I noted there that though Gilbert sees the idea of a joint commitment as a primitive, she does want to make substantive and informative claims about joint commitments. She wants to offer claims that will to some extent help us to understand what this primitive phenomenon is. As we have seen, one of these substantive claims is that mutual obligations inhere in joint commitments; and this is a claim I have examined at some length. But it might also have seemed that a second substantive claim was that joint commitments induce plural subjects. However, for this

[258] The correspondence is December, 2008. The quoted passage is in A Theory of Political Obligation (Oxford: Oxford University Press, 2006) at 144-5.

to add something substantive to the theory, we cannot say that talk here of a plural subject is just a facon de parler.

As I noted earlier, one question we can have about the very idea of joint commitment is: in what does the jointness of the commitment consist? What makes it our joint commitment and not merely a concatenation of commitments of each? The basic thesis understands the joint-ness or shared-ness of shared intention as consisting in a web of inter-relations among relevant plan states with common social contents. In contrast, Gilbert sees joint commitment as a primitive relational phenomenon. Still, without seeking a reduction of joint commitment to other things, we can ask whether anything substantive and informative can be said about the nature of such joint-ness. And here it might have seemed that the appeal to the idea of a plural subject could be a part of Gilbert's answer to this question. But on the second – privileged - interpretation of Gilbert's talk of a plural subject, this cannot be so. This might exert pressure to return to the first, more robust interpretation of this talk of a plural subject; but I have argued against the claim that plural subjects, so interpreted, are present in all cases of modest sociality.

In discussing a view of shared action that is broadly in the spirit of my view, Philip Pettit and David Schweikard observe that this view "fails to point us to a single collective subject that is causally responsible for the action"; but they also go on to say that they

> see no metaphysical reason why a joint intentional action has to be the product of a single agent or a single state of intending. …
>
> [we do not need] the joint construction of a novel center of intentional attitude and action.[259]

I am here agreeing with Pettit and Schweikard that we need not suppose that in modest sociality there is a "novel center of intentional attitude" – a subject, in a strict sense, of the shared intention. But I think that Petersson is right to point us to the idea that there can nevertheless be, in a weak sense, a "single agent" -- though not "a single collective subject" – that is, in the words of Pettit and Schweikard, "causally responsible for the action". This agent is the internally structured group – where that group consists

[259] Philip Pettit and David Schweikard, "Joint Actions and Group Agents," 30. As noted, Pettit and Schweikard interpret Gilbert as asserting that there is such a "novel center" in joint intentional action.

of the appropriately structured collection of the participants (and where the basic thesis is a thesis about what in the present case that structure is). But, given the moderate holism of subject-hood, in cases of modest sociality we should not expect that this group is "a single collective subject" or "a novel center of intentional attitude". In modest sociality, then, group agents of a shared intentional activity – agents that are "causally responsible for the action" – are not in general subjects of the shared attitudes that help make it true that there is shared intentional activity.[260]

As Joshua Cohen has emphasized (in conversation), drawing on ideas of John Rawls, we need to be careful when we try to extend our model of individual agency directly to a model of shared agency. And what we have seen is that this thought applies to the way we think of the connection between agent-hood and subject-hood. In the case of individual intentional agency it is plausible to agree with Davidson that the agent of the intentional activity is the subject of a more or less holistic web of attitudes, some of which are part of the explanation of that activity. But what we have seen is that when we turn to modest sociality, a correspondingly tight connection between agency and subject-hood is less plausible.

Return now to the purported own-action condition on intention. When I first discussed that condition in Chapter 1 I said that it seemed initially that appeal to our shared intention to J – in contrast with my intention that we J -- satisfies that own-action condition. But now we can see that matters are actually more complicated. Talk of our intention in favor of our action is, to be sure, in the spirit of the own-action condition. But we have now seen that in modest sociality this talk of our intention need not treat us as the subject of the shared intention, in the strict sense of a subject of a broadly holistic mental web. There is, to be sure, a match in the dual use of the first person plural in talk of our intention that we act. But, strictly speaking, such talk need not pick out a subject of the shared intention who is one and the same as the agent of the intended action, since such talk need not pick out a subject of the shared intention. Still, if all

[260] As the quotations above from Pettit and Schweikard indicate, they are assuming a tight connection between group agent-hood and group subject-hood. And the assumption that a group agent is a group subject is also central to List and Pettit, Group Agency. But I do not see reason to think that these philosophers would need to reject the claim I am making, in following Petersson, that there is a weaker kind of group agency that does not require subject-hood though it does require appropriate internal organization of the group.

goes well there will be a structured collection of participants that is the agent of the shared activity and also is the locus of those inter-connected attitudes that constitute the shared intention in favor of that activity. So something in the spirit of the own-action condition is true about our shared intention, and there remains a contrast between talk of our intention to J and my intention that we J.

Chapter Eight: Shared Deliberation and Shared Policies of Acceptance[261]

1. Two aspects of shared deliberation

A central claim is that the rational functioning characteristic of modest sociality emerges from the rational functioning of the individualistic intention-structures described by the basic thesis. In my partial defense of this claim in Chapter 4, I put aside one aspect of the rational functioning of shared intention, namely its role not only in providing a framework for bargaining, but also in providing a framework for shared deliberation. I now want to fill in this lacuna.

Let me begin by highlighting two central features of shared deliberation. The first is that shared deliberation is something we do together: it is not simply a concatenation of the reasoning of each, in a context of common knowledge, about what it would be best for us to do.[262] Shared deliberation is itself a distinctive form of modest sociality.

This is a feature that shared deliberation has in common with shared bargaining: in each case we are thinking together. It is a second feature that distinguishes shared deliberation from ordinary bargaining.[263] In ordinary bargaining we each bring to bear considerations that matter to each of us without assuming that the very same considerations directly matter to the others. In shared deliberation, in contrast, we reason together in a way that involves shared commitments to treating certain considerations as mattering in a certain way.[264] We can call these shared commitments to weights.[265]

To make progress, then, we need to clarify these two central aspects of shared deliberation: the way in which shared deliberation is itself a form of modest sociality, and the nature and role of shared commitments to weights.

---

[261] Elements of this chapter are drawn from my "Shared Valuing and Frameworks for Practical Reasoning".
[262] This contrasts with "team reasoning" as understood by Gold and Sugden. See Chapter 5.
[263] As noted in Chapter 1, the distinction between bargaining and shared deliberation is in the spirit of a similar distinction made by Andrea Westlund in her "Deciding Together," Philosophers' Imprint vol. 9, No. 10 (2009).
[264] These considerations may include procedural constraints, such as a need for consensus.
[265] The metaphor of weights may not be fully apt for certain kinds of side-constraints, but we can safely put this complexity to one side here.

Let's begin with the way in which shared deliberation is a form of modest sociality. A first step is to apply a central idea of the basic thesis and say that in shared deliberation we reason together by way of a shared intention to deliberate together. Indeed, as we have noted, in shared deliberation we reason together in a specific way, namely: we bring to bear shared commitments to weights. So we can say that in shared deliberation we reason together by way of bringing to bear shared commitments to weights, where this entire activity is guided by a shared intention to deliberate together by way of such shared commitments.

While this is true, however, it is limited in its explanatory power, since the very idea of shared deliberation by way of shared commitments to weights appears in the content of the cited shared intention. So let's proceed in two steps. First, let's take as given the idea of shared commitments to weights and see if, with this idea in hand, we can develop these suggestions into an account of shared deliberation that is not unacceptably circular. Such an account in hand, we can then try to explain shared commitment to weights in more detail.

In taking the first step we can apply our general strategy for responding to such threatened circularities: we seek a description of our joint activity that, while substantive, is neutral with respect to shared intentionality. So consider joint activity in which each gives weight to R in his own thinking and in arguments offered in public, in a context in which others do so as well; each is prepared to criticize himself and others for failing to accord such weight in such personal and public thinking; each is prepared to offer corresponding advice; and all this is out in the open. This is a complex of interrelated intentional activities characterized in a way that seems to be neutral with respect to shared intentionality. Call this joint reasoning.

Can we say that shared deliberation is joint reasoning guided by a shared intention in favor of joint reasoning? No we cannot. This is because in saying this we would not be ensuring a role in shared deliberation not just of each participant's commitments to relevant weights, but of shared commitments to relevant weights. So let us say instead that shared deliberation is joint reasoning that is guided in particular by shared commitments to weights, where this entire activity is guided by a shared intention in favor of joint reasoning that is guided by shared commitments to weights.

But what is a shared commitment to weights? My proposal will be to understand a shared commitment to weights as a special kind of shared intention, as that is understood by the basic thesis. The rational functioning of such a shared intention about weights will emerge in familiar ways from the rational functioning of the individualistic attitudes that constitute that shared intention.

2. Shared commitments to weights

Begin by considering some examples of a shared commitment to weights. Perhaps we share a commitment within our shared painting of the house to our giving substantial weight to environmental concerns as we decide what paints to use, and how to dispose of various materials. Or perhaps we are engaged in a shared intentional activity of building a house and we deliberate together about sub-plans in a way that brings to bear our shared commitment to giving weight to certain standards of earthquake safety. An admissions committee might have a shared commitment to its giving weight to legacy considerations in its admissions decisions. A scientific research team might have a shared commitment to its giving weight to earning lucrative patents in its decisions about the direction of its research. An academic department might have a shared commitment to its giving weight to collegiality in making faculty appointments. A political group might have a shared commitment to its giving weight to promoting a progressive tax structure. A committee might have a shared commitment to its giving weight to consensus in its decision-making.[266]

Let's try to lay the groundwork for a systematic treatment of this phenomenon. Consider a Philosophy department that has a shared commitment to its giving weight to issues of sub-field in its searches for new faculty and in its associated deliberations.[267] This is a matter on which different departments may reasonably diverge. One department may have a shared commitment not to give weight to such issues; another

[266] As these examples indicate, this phenomenon sometimes is embedded within larger institutional structures of the sort that I have been trying to put to one side for present purposes. So we need to be careful that we identify features of this phenomenon that do not essentially depend on such an institutional embedding. My hope is that the discussion to follow succeeds in this respect.

[267] For discussions of related examples see J. David Velleman, "How To Share an Intention," as reprinted in his The Possibility of Practical Reason (Oxford: Oxford University Press, 2000), 200-220, and Christopher Kutz, "The Judicial Community," Philosophical Issues 11 (2001): 442-469. In keeping with caution about embeddings in institutions, I am here ignoring the authority of the chairperson of the department (as do most philosophers).

may have a shared commitment to giving great weight to whether candidates are in a previously designated sub-field. In each case the shared commitments are, we may suppose, grounded to some extent in judgments about what makes a good department, as well as in other shared commitments (e.g., about who gets to vote on such matters). It is possible that the members of the department agree about which shared commitment would be (prior to any departmental commitment to a policy) best, and that this is why the department has that commitment. But such shared commitments may go beyond such judgments about value, and do not require agreement in those judgments. In participating in such a shared commitment one need not suppose that it is the best such shared commitment. One may think there is no single best; or one may think that a different shared commitment would be best. Indeed, each member of the department may have a different view of what the best shared commitment would be, and yet arrive at a shared commitment that no one sees -- at least prior to the shared commitment -- as best. Such shared commitments nevertheless help structure shared deliberation and planning.

Again – and returning to an earlier example -- different admission committees may have different shared commitments about the justifying weight to be given to legacy considerations in deliberations about undergraduate admissions. One may work in the admissions office and be committed, as a member of that office, to giving a certain weight to such considerations without thinking this is the best such shared commitment.

In some cases particular shared commitments concerning what to treat as having weight in certain contexts of shared activity are more or less definitive of the group whose shared commitments they are. If you are going to be a member of a certain scientific research group, you may need to participate in a shared commitment to giving weight in scientific debate to certain kinds of evidence, but not to whether or not the person offering the evidence is your friend. Central to certain groups may be a shared commitment to treating conformity to particular religious texts or traditions or rituals as a justifying reason for action. Again, a club may have a shared commitment to its giving justifying weight to, say, religious affiliation, or race, in its deliberations about membership; and that may be why you do not want to be a member of that club.

As these examples suggest, such shared commitments about weights in shared deliberation will normally be part of a larger package of shared commitments, one that includes shared intentions in favor of associated shared activities. A group might have both a shared intention to worship together each Saturday, and a shared commitment to treat associated rituals and traditions as providing justifying reasons for action on those occasions of shared worship. And a scientific research team might have a shared intention to engage in a certain line of research, together with shared commitments about what is to count as a justifying consideration in the group's associated deliberations about what and when to publish. In each case participation in the relevant shared intentional activities of religious worship or scientific research involves as well participation in related shared commitments about what to treat as a justifying reason in the context of those shared activities. Indeed, as Joseph Shieber once emphasized (in conversation), such shared commitments about what to treat as a justifying reason can tie a group together even in the face of some divergence of view concerning particular plans of action. And, as we have seen, such shared commitments about what to treat as a reason can help frame shared deliberation even if there is reasonable disagreement about which such shared commitment would be best.

3. Shared policies about weights

How might we model such shared commitments to weights? We might try saying that such shared commitments are a matter of converging judgments of value in a context of common knowledge: our shared commitment to our giving weight to X consists in our each judging that X is valuable, in a context of common knowledge of these judgments. The problem, however, is that, as has already been noted, sameness of value judgment, in a context of common knowledge, seems neither sufficient nor necessary for a corresponding shared commitment to weights in shared deliberation. Perhaps each of us thinks earthquake safety in construction is a good thing, and these judgments are common knowledge; yet we still do not have a shared commitment to give weight to earthquake safety in our shared deliberations concerning our construction project. Perhaps some of us resist such a commitment because, though earthquake safety is a good thing, it would stand in the way of higher profits, or speed of construction. Again, it seems to be one thing for each of us to think that consensus is a

good thing and for these judgments to be common knowledge, another for us to share a commitment to our giving the pursuit of consensus weight in our deliberative activities. Even after it is common knowledge that we each judge that consensus is of value, there seems a further step that is needed for us to have a shared commitment to our giving weight to consensus in our shared deliberation. And some groups might not take that step. Further, even if we do each judge that earthquake safety – or consensus -- is a good, and do go on to a shared commitment to relevant weights in shared deliberation, it seems that these weights need not correspond to a common weight in our value judgments. After all, there might be no such common weights in our value judgments. Finally, it seems that we could establish a shared commitment to, for example, our giving weight to collegiality in our hiring decisions, or to legacies in admissions decisions, even if some of us are skeptical about the associated values but participate in the shared commitment as part of a social compromise. Such a shared commitment does not, then, require agreement in value judgment.

A shared commitment to weights seems then to be a different phenomenon than that of a convergence of value judgment. Sharing a commitment to certain weights seems closer to a kind of shared intention than to a common value judgment. This suggests that our shared commitment to weights is better modeled as a shared intention to our giving weights to certain considerations in relevant shared deliberation. Or better – to capture a characteristic generality -- as a shared policy about weights in relevant shared deliberation. (Recall that policies are intentions with relevant generality.) Here again, what is central to our modest sociality is the sharing of intentions and plans, and not agreement in belief or judgment: this is another aspect of the primacy of intention for modest sociality.

I noted in Chapter 1 that an individual planning agent may have policies about weights for individual deliberation, policies that may well go beyond or in other ways diverge from her associated judgments of value. And now the suggestion is that we understand shared commitments to weights as a shared version of such policies about weights.

This suggests that we apply our approach to intention sharing to the case of shared commitments in favor of our giving relevant weights in shared deliberation. We identify such shared commitments about weights with shared policies about weights in shared deliberation. And we apply the basic thesis in understanding such shared policies. On this approach, a shared commitment to give weight to R is a shared policy to give weight to R in relevant shared deliberation. And this shared policy consists, roughly, in interlocking and interdependent general intentions, on the part of each, in favor of our giving weight to R in relevant shared deliberation -- where all this (as well as associated beliefs about interdependence and efficacy) is in a context of common knowledge. And the rational functioning of such a shared policy will emerge, in now-familiar ways, from that of its individualistic constituents.

Such shared policies about weights will have complex relations to the value judgments of the participants. It will normally be true that each participant's policy contribution to the shared policy about weights will be to some extent grounded in her associated judgments of value. Nevertheless, the proposal is that the shared commitment to weights does not consist in these value judgments, but rather in the shared policy about weights. Such shared policies provide relevant normative and/or evaluative premises for shared deliberation. These shared policies will help structure relevant shared deliberation, and will engage associated rational pressures of consistency, agglomeration, coherence, and stability. And since such shared policies will normally be limited to specific contexts of sharing – a particular construction project, say; or a particular admissions process – they will be context-relative in a way in which value judgments are not.[268]

[268] Here I am drawing on ideas in my "Practical Reasoning and Acceptance in a Context," (1992) as reprinted in Faces of Intention. For a useful discussion of related ideas see Anthonie Meijers, "Collective Agents and Cognitive Attitudes," Protosociology 16 (2002): 70-86. Meijers argues that all cases of "collective belief" that are not merely a summative matter are ones of context-relative collective acceptance. (For a related view see K. Brad Wray, "Collective Belief and Acceptance," Synthese 129 (2001): 319-333. And for an alternative view see Margaret Gilbert, "Belief and Acceptance as Features of Groups," Protosociology 16 (2002): 35-69.) In contrast with Meijers, my claim is only the weaker claim that there is a phenomenon of shared commitments to weights that is to be identified with a shared policy about weights, a shared policy that is context-relative in the way I have indicated; and below, in section 7, I go on to a more general idea of a shared policy of context-relative acceptance.

Recall that in Gricean creature construction we try to see each further step in the construction as building on the resources of the prior step in response to some recognizable pressure or problem at that prior stage. And that is how we can see the step to shared policies about weights. Roughly speaking, the basic move to shared intentions to act together is a response to a need for increased social coordination and unity of thought and action. And here I have sought – in the form of the basic thesis -- a conservative construction of such shared intentions. The further move that we are now considering – a move to shared policies about weights -- can be seen as in part a response to further problems that emerge in the effort to achieve, by way of shared thinking, meshing sub-plans for such shared intentions. These problems may be acute, given potential divergence in relevant views about the right and the good that are held by the various participants. And here, again, I seek a conservative construction, by way of the basic thesis. We build such shared policies about weights out of the materials of policies of the individuals about weights, where those policies concern, in particular, the context of relevant shared deliberation. This is to extend the architecture of the basic thesis about shared intention to shared policies about weights.

Shared policies about weights in shared deliberation can, in complex ways, recognize and build on relevant evaluative agreement, if such there be. But such shared policies also frequently go on to further shared commitments in order to make shared deliberation more determinate and more likely to support needs for mesh in our modest sociality. In this way such shared policies about weights can help make possible complex forms of shared activity in the face of background differences in value judgment.[269]

It is important that these shared policies about weights can support such determinateness and social mesh without requiring that the participants respond to these social pressures by changing their evaluative beliefs or judgments. After all, changing one's beliefs or judgments in response to such practical pressures would be in tension with the nature of belief or judgment: we do not normally see rational belief and

[269] In "Shared Valuing and Frameworks for Practical Reasoning" I argue that it is reasonable to see such shared policies about weights as, in particular, a kind of shared valuing. While I continue to see this proposal as theoretically attractive, we can safely put it to one side for present purposes.

judgment as directly responsive to such practical pressures. These are, as is said, the wrong kind of reasons for belief and judgment. But there is not the same problem in seeing policies about weights as a response to such practical pressures.[270]

For similar reasons I think it would be misleading to say that these shared policies about weights constitute the group's shared beliefs or judgments about values or reasons. Such shared policies are substantially – and reasonably -- shaped by practical concerns in ways in which we do not normally see as legitimate for belief or judgment.

When one uses a certain evaluative or normative premise in deliberation, one accepts that premise in that deliberative context.[271] Shared policies about weights are, then, shared commitments in favor of our accepting certain evaluative or normative premises in certain contexts of shared deliberation. And a shared commitment to such shared context-relative acceptance of such premises can be a reasonable response to broad practical pressures of sociality.

A shared policy to give weight to R differs importantly from a case in which it is simply true that each in fact gives, or intends to give, weight to R, and this is common knowledge.[272] When there is this shared policy each not only takes note of the fact that each gives weight to R. Each (interdependently) intends that each give weight to R, and that this proceed by way of each participant's so intending.

Our shared policy to give weight to R involves policies on the part of each in favor of our giving weight to R in shared deliberation. To avoid an unacceptable circularity, the appeal in this content to shared deliberation is in basic cases to be

[270] Wray makes a closely related point in his "Collective Belief and Acceptance," at 325.
[271] See, again, my "Practical Reasoning and Acceptance in a Context," in Faces of Intention; and my "Three Theories of Self-governance" in Structures of Agency, at 240. And see Robert Stalnaker, Inquiry (Cambridge: MIT Press, 1984) chap. 5.
[272] Something close to a version of this contrasting case plays a central role in the model of team reasoning proposed by Gold and Sugden. See their "Schema 4" in their "Collective Intentions and Team Agency," at p. 126. Their "team pay-off function" is a matter of a common pay-off function in a context of common knowledge. In contrast, a shared policy about weights involves a shared intention in favor of the common treatment of certain considerations as supporting certain decisions. In commenting on my "Shared Valuing and Frameworks for Practical Reasoning" Gold and Sugden remark that "a shared policy that says what considerations are to be given justifying significance … would be the analogue of the team pay-off function." (137) But this ignores the difference just noted between such a shared policy and a common pay-off function in a context of common knowledge.

captured by an appeal to joint reasoning understood as a web of inter-related intentional activities. Recall that this web includes actual and potential advice or criticism of self and others, all of which is out in the open. And the shared policies about weights provide a ground for that advice or criticism. Since the relevant policy is a shared policy, each can say, both to herself and to others, that failure to give appropriate weight to R is inconsistent with a policy that is held both by the person offering the advice or criticism and by the person to whom the advice or criticism is directed. (In self-advice or self-criticism, it is the same person.) So -- to draw from H.L.A. Hart -- in such cases each treats giving weight to R as a "common pubic standard".[273] And the presence of such a common public standard helps explain the tendency to conformity with that standard.

Shared deliberation, then, is joint reasoning guided by relevant shared policies about weights, where this entire activity is guided by a shared intention in its favor. In the absence of relevant shared policies about weights the parties may engage in modest sociality and related bargaining; but they will not be in a position to engage in shared deliberation. A corollary – anticipated earlier -- is that insofar as the kind of sociality at hand is solely that of a social network, there is no guarantee that there are sufficient resources for shared deliberation.

This account shows how various kinds of shared intentions – understood in terms of the basic thesis -- will tend to frame shared deliberation. Shared deliberation will typically take place within the framework of an overarching shared intention and aim at arriving at meshing sub-plans. In the involved shared deliberation, shared policies about weights will play central roles. Such shared policies about weights are intention-structures of the sort described by the basic thesis. And this completes the defense of the claim that modest sociality emerges from the forms of inter-connected planning agency highlighted by the basic thesis.

4. Interdependence in policies about weights

Shared commitments to weights are shared policies about weights. A central aspect of such sharing, according to the basic thesis, is interdependence in persistence

[273] H.L.A. Hart, The Concept of Law at 116.

between the relevant policies of each. Let us reflect further on the nature of this interdependence.

The first point is that in many cases such interdependence will be a result of a recognized need, in the pursuit of relevant social unity, to fix on a group policy concerning weights despite divergence in relevant value judgments of the individuals. We each are settled on this policy about weights, in the face of divergence in value judgment, in part for the reason that the others are settled on this policy as well. There will, however, also be cases of such interdependence in the presence of agreement in relevant value judgment. Perhaps all the members of an admissions committee agree, for similar reasons, that a particular policy about legacy considerations would be best. Nevertheless, it may also be true that, given their recognition of the need for coordination in such matters, they are each committed to the policy in part also for the reason that others are.

Recall that in Chapter 3 we distinguished between extensive interdependence and limited interdependence between our intentions that we J. The dependence of my relevant intentions on yours is extensive if it includes not only dependence of my intention that we J on your relevant intentions, but also associated dependencies on your intentions of my intention to do what would be my part in J. The dependence of my relevant intentions on yours is limited if it includes only dependence of my intention that we J on your relevant intentions. And in each case we allowed for either reason-for or enabling interdependence.

Our interest now is in the case in which J is: our giving weight to R in relevant shared deliberation. (Here I help myself to the idea of our shared intentional activity of giving such weight, given our strategy for avoiding circularity.) Simplifying a bit,[274] we can consider four main kinds of interdependence between such policies about weights:

	Enabling dependency	Reason-for dependency
Interdependence between intentions that we give weight to R in relevant shared deliberation	A	B

[274] The matrix to follow limits attention to versions of boxes I and IV in the matrix in Chapter 3 section 4.

Interdependence between intentions to perform what would be one's own part in such shared deliberation	C	D

Cases of extensive interdependence will involve interdependence on both the top and bottom rows; cases of limited interdependence will be limited to the top row.

Begin by reflecting on the top row: interdependence between each of our intentions in favor of our giving such weight in shared deliberation. Such interdependence will frequently involve reason-for interdependence, and so be in box B. Indeed, this is the kind of interdependence illustrated in the case just noted in which we each have the policy in favor of our giving certain weights, in the face of divergence in value judgment, in part for the reason that the other has this policy. But sometimes the interdependence might primarily be only an enabling interdependence in which each sees the other's policy in favor of their giving weight to a certain consideration not as a reason for favoring their giving such weight in shared deliberation, but only as an enabling condition for their deliberating in this way. Perhaps you and I each have a policy that supports our giving weight, in the context of our joint research project, to seeking out related published works and acknowledging them in our publications. We each recognize that if the other did not have this policy it would not be reasonable to suppose that we will indeed deliberate in this way. And we recognize that our shared project might then break down. But neither of us sees the other person's policy as his reason for favoring our giving weight to such scholarly practices. Each person's reason for favoring this practice is, quite simply, the recognized value of the practice, not the other's policy in favor of that practice; though each nevertheless intends that they reason in this way (where this involves the contributions of both). So this is a case of enabling, but not reason-for inter-dependency on the top row: a case located in box A.[275]

[275] This case is modeled on Scott Shapiro's example of the "fundamentalists" in his Legality, at 109. Shapiro sees such examples as showing that there can be important forms of sociality despite a kind of independence between the relevant commitments of each of the participants. And Shapiro sees this fact as posing a challenge to H.L.A. Hart's account of social rules. However, what we can learn from the more nuanced account of interdependence that I am sketching here is that though there is the absence of

That said, it is also true that even in such a case there will be pressure in the direction of at least minimally reason-for interdependence. This pressure comes from the structure of the targeted shared activity: in this case, the joint research project. After all, it is required that there be intentions that this activity proceed by way of meshing sub-plans. And part of the way the activity proceeds is by way of relevant, shared deliberation. So each participant is committed to there being meshing sub-plans with respect to that deliberation. So each participant is committed to there being an appropriate mesh in the relevant policies about what to give weight to in that deliberation. So each is under some rational pressure to adjust her policies about such weights in order to mesh with her partners' corresponding policies. So when there is such mesh -- as there is in our case in which each has the same policy about the weight to be given to certain kinds of scholarship -- each will be under some rational pressure in favor of her continued contribution to that mesh, rational pressure that derives from each person's commitment to mesh, together with facts about the other's policy. Granted, it remains possible that this rational pressure is not part of the person's reasons for her relevant policy. So reason-for interdependence is not ineluctable in such cases. But it also seems that, given the intentional structure of the shared activity and associated shared deliberation, there will normally be at least minimal reason-for interdependence.

Turn now to the distinction between extensive and limited dependency. Extensive dependency is both top-row and bottom-row; limited dependency is solely top row. There is limited interdependence in policies about weights when (a) each person's policy in favor of the group's giving R weight is dependent on the other's corresponding policy, but it is not true that (b) each person's policy in favor of doing what would be his own part in such giving of weights is dependent on the other's corresponding policy in favor of doing what would be his own part.

---

relevant reason-for interdependence among Shapiro's fundamentalists, there may well still be relevant enabling interdependence. Further, and as I go on to discuss in the main text, there will also remain certain kinds of rational pressures in the direction of relevant reason-for interdependence. So such fundamentalists can engage in shared deliberation. I suspect that these observations can help us defend a version of Hart's theory from this specific criticism from Shapiro, but I will not pursue this issue here. (I do pursue it in "Law and Common Public Standards," unpublished MS.)

What would it be like for (a) to be true and yet (b) false? Well, a policy to do what would be one's own part in such giving of weights would favor activities like trying to argue to the group in favor of its giving weight to R. And it seems that one's policies in favor of the group's giving weight to R may be interdependent even if such policies of each about what would be her own part in such giving of weights are not interdependent. Perhaps Jones and Smith each have policies in favor of the group's giving weight to R, and these policies are inter-dependent in the sense of (a). But perhaps Jones would argue for and otherwise try to induce this commitment in the group by way of his own personal activities even if Smith himself had no such policy of personal support. So in this case (b) is false. This would then be a case of limited interdependency.

In my discussion in Chapter 3, I concluded that in modest sociality the interdependence could be either limited or extensive, and could be reason-for or merely enabling. And it seems to me that a similar, pluralistic approach is appropriate for the case of shared commitments to weights. Frequently the interdependence will be extensive and reason-for. However, cases of limited and/or only enabling interdependence seem possible. And our theory can allow for cases in which what frames the shared deliberation involves such somewhat weaker forms of interdependence.

5. Partiality and depth of shared policies about weights

Shared commitments to weights are shared policies about weights in relevant shared deliberation. These shared policies will normally be to some extent responsive to relevant judgments of value on the part of the various participants. Nevertheless, these shared policies do not require convergence in supporting value judgments.

This is another aspect of the pervasiveness of partiality in sociality. I have emphasized that shared intention in favor of shared action need not involve commonality of reasons for participating in the sharing: a shared intention can involve a convergence on the shared activity that is grounded in diverging background reasons. We now see how this point extends to shared commitments to weights. Those on a college admissions committee might have a shared policy of giving weight to legacy considerations in relevant shared deliberation. Perhaps some participate in this because

they think giving such weight in admissions decisions is an effective fund-raising tool; whereas others participate because they think their institution has made an implicit promise to its alumni to provide this benefit to their children. Members of the committee participate for different reasons: their sharing is in this way partial. But their shared policy about weights nevertheless establishes a common and interlocking – albeit, partial and somewhat shallow – framework for their shared deliberation.

As in the case of shared intention concerning first-order activity, I think it is a virtue of our theory of shared policies about weights in shared deliberation that it helps us model such partial but substantial social unity in the face of different background reasons for participation. Much of our sociality is partial in this way, given the pressures for shared agency in the face of such differences, pressures and differences that are characteristic of, in particular, a pluralistic, liberal political culture. We manage to reason together in the pursuit of shared projects despite significant background differences of view. We can many times reason together in these ways without seeking and/or arriving at a consistent, overall set of social judgments about the right and the good. So the pervasiveness of partiality in the social extends in significant ways to the grounds for which different agents participate in shared commitments to weights. Our model of shared commitments to weights as shared policies about weights aims to provide in a clear way for these important phenomena.

The idea then is that, in response to practical pressures, we might reach a more extensive agreement about weights for our shared deliberation. This would be puzzling if we thought that what we would be agreeing to is a belief or judgment about which weights are the correct weights. We do not normally think of belief or judgment as rationally responding to such practical pressures. But in agreeing to participate in a shared policy about weights I need not be agreeing to change my beliefs or judgments about these weights. I need only agree to be guided by a relevant policy in certain social contexts.

This is important since convergence in belief or judgment is frequently quite difficult and unlikely to be achieved. So we want to model forms of shared agency that

need not depend on, and do not include a demand in favor of such convergence.[276] And this is, again, an aspect of the primacy of intention in our sociality.

The theory also leaves room for efforts on our part – when we think they are called for -- to reach a deeper and more extensive agreement concerning the background rationale for such sharing. Such efforts may involve further agreements concerning what is to count as justifying within relevant shared deliberation. These efforts may themselves be shared intentional activities that are guided by our shared intention to achieve such further depth of agreement -- in response, perhaps, to concerns about the determinateness of our shared deliberations in hard cases.[277]

After all, as Sheanna Shiffrin has noted in conversation, if our shared policy about weights also involves a common, substantive background rationale (over and above general reasons for coordination) we will probably be in a better position to respond to complexities about how to proceed in hard cases than we are when the reasons for which we each participate differ. If, for example, our shared policy of giving weight to legacy considerations is grounded in a shared policy of giving weight to creating loyalty among potential donors (and not grounded in a supposed promissory obligation to alumni), we would probably be in a better position to resolve certain disputes in hard cases. In a case involving conflict between legacy considerations and other considerations that we value – ethnic or economic diversity in the student population, for example – we might be able to settle disputes in part by appeal to a shared policy about the background reasons for the shared policy in favor of legacies. In contrast, if our common framework is a thin one that consists entirely of our shared policy of giving weight to legacies – where we each have different background reasons for our participation in this common framework (over and above general reasons for coordination) -- we will not have available this potential contribution to conflict resolution.

---

[276] This is related to a central theme in John Rawls's political philosophy. See his discussion of "the fact of pluralism" in his "The Domain of the Political and Overlapping Consensus," in John Rawls: Collected Papers (Harvard University Press, 1999): 473-96. Related issues are also discussed by Blain Neufeld in his "The Idea of A Civic People," unpublished manuscript.

[277] For talk of "hard cases" see Ronald Dworkin, "Hard Cases," as reprinted in his Taking Rights Seriously (Cambridge: Harvard University Press, 1977), 81-130.

Of course, in a particular case it may not be possible for us to go beyond such a thin common framework. Indeed, there may in certain cases be substantive pressures for us not to go beyond such a relatively shallow commonality, but rather to retain a kind of neutrality with respect to these differences. These may include certain cases of what Cass Sunstein calls "incompletely theorized agreements".[278] However, Shiffrin's point is that if this is the kind of shared framework in which we are participating, then there is a significant risk that we will have difficulties in applying the shared framework to hard cases. And that might be why we aim to reach further agreement about the substantive background for our shared policy about the weight of legacy considerations.

6. Shared Policies of Acceptance

Shared policies about weights involve interlocking and interdependent policies on the part of each in favor of giving weight to certain considerations in shared deliberation. These interlocking and interdependent policies provide relevant normative or evaluative premises for a specific context of shared deliberation. The participants need not believe these premises; they need only be committed to accepting them in the context of relevant shared deliberation. Such interlocking and interdependent policies in favor of such acceptance-in-a-context will sometimes be a reasonable response to a recognized need for determinateness of shared normative or evaluative background in the face of divergence in normative or evaluative judgment.[279]

And now the point to note is that this solution generalizes: it applies quite generally to cases of sociality in which we need determinateness of shared background despite divergence in judgment or belief. We have just now been focusing on the case in which the divergence is in value judgment or the like, and the social solution has been a shared policy about weights. But there can also be divergence in relevant beliefs about matters that are not explicitly evaluative -– about, for example, what is possible and what would be effective. Here too shared deliberation and shared agency frequently require a determinate shared background. And we can sometimes respond

[278] Cass Sunstein, "Incompletely Theorized Agreements" Harvard Law Review 108 (1994-95): 1733-1772.

[279] And similarly with respect to under-determination of a shared perspective that is due not to the divergence but to the partiality and incompleteness of the evaluative judgments of the participants. Here, however, I focus on cases of divergence.

to this need for determinateness of shared background by way of a shared commitment to accept certain propositions in our shared deliberations, where this need not involve a convergence in our beliefs about these matters. And, again, we can model such a shared commitment as a shared policy in favor of certain forms of context-relative acceptance.

I noted in chapter 1 that shared intentions and plans will be held against a cognitive background that concerns, roughly, what is possible and what is effective. To simplify my discussion I initially assumed that the participants have the same beliefs about these matters. And I appealed to this simplifying assumption in expressing, in a rough way, the basic social norms on shared intention of social consistency and coherence. We are now in a position to deepen this aspect of the theory.

The basic point is that sometimes in shared agency we can, in the face of differences of belief concerning relevant matters of possibility or effectiveness, arrive at a shared policy to accept a relevant proposition in the context of our shared deliberations. Perhaps we are engaged in a shared project of building a house together, but we disagree about the likely costs of certain elements of the project. We might agree to take certain estimates of the costs as given for our shared deliberation – estimates that perhaps do not match anyone's actual beliefs. We thereby come to have a shared policy in favor of our accepting these estimates in the context of our relevant shared deliberation.[280]

The target of this shared policy of context-relative acceptance is not an evaluative or normative premise, but a factual premise. Nevertheless, it seems plausible to understand this shared policy of acceptance in terms of ideas we used to understand shared policies about weights. We appeal to the idea that we each intend that we take p as a given premise in relevant shared deliberation and that this proceed by way of each of our so intending. We appeal to the idea that these policy-like intentions of each, concerning what we are to take as given in our shared deliberation, are appropriately interlocking and interdependent. And so on.

Shared policies about weights are a special case of such shared policies of context-relative acceptance. Shared policies about weights and, more generally, shared

[280] For a version of this example see my "Practical Reasoning and Acceptance in a Context," at 24-25.

policies of acceptance, play a framework-providing role in shared deliberation. We understand these framework-providing shared policies of acceptance in shared deliberation by extending the architecture of the basic thesis. And by understanding these shared commitments as forms of shared intention/shared policy – rather than of common judgment or belief – we make better sense of the various practical pressures on these shared attitudes.

A shared policy of acceptance that p in the context of relevant shared deliberation involves interlocking and interdependent policies of each in favor of accepting that p in relevant contexts. And when such a shared policy operates effectively there will indeed be shared context-relative acceptance of the premise that p. A shared policy of acceptance that p is not just a matter of common knowledge or common belief of the fact that each accepts that p in relevant contexts. When there is such a shared policy of acceptance each not only knows or believes that each accepts that p. When there is such a shared policy of acceptance each intends that each accept that p in relevant contexts and that this proceed by way of the cited intentions of each and their meshing sub-plans. The common acceptance of p is something each intends, rather than simply takes note of. As a result there will normally be an intention-supported stability of the pattern of common acceptance that p, an intention-supported stability that need not be ensured simply by common knowledge or common belief.[281]

So we have, at the individual level, intentions and plans concerning individual activities, and policies concerning weights; and we have, at the shared level, shared intentions concerning shared activities and shared policies concerning weights and, more generally, shared policies concerning acceptances in relevant shared contexts. These various intention-like attitudes play fundamental roles in the cross-temporal and

[281] Robert Stalnaker offers a model of what he calls (following Paul Grice) "common ground" for a conversation, a model that appeals to common belief in a pattern of acceptances. As he says: "It is common ground that phi in a group if all members accept (for the purpose of the conversation) that phi, and all believe that all accept that phi, and all believe that all believe that all accept that phi, etc." (716) My remarks in this paragraph are intended to point to an important difference between shared policies of acceptance and a common ground in Stalnaker's sense. I do not say that Stalnaker's account of common ground is not apt for his purposes of analyzing important features of conversation, including presupposition (which is one of his main concerns). My purpose in appealing to shared policies of acceptance is different from Stalnaker's purpose in analyzing conversation. My purpose is to provide a helpful model of shared deliberation, and not all conversation is shared deliberation. See Robert Stalnaker, "Common Ground," Linguistics and Philosophy 25 (2002): 701-721, at p. 716. Thanks to Joshua Armstrong for pointing me to this essay.

social organization of thought and action, individual and shared. They do this in part by responding to pressures for determinateness in thought and action in the face of under-determination by relevant judgments and, in the social case, divergence in such judgments.

Return now to the norms on shared intentions of social consistency and social means-end coherence. As noted, when I first described these norms I made the simplifying assumption of sameness of belief concerning what is possible and effective. But we have noted that sometimes there is divergence in these beliefs of the participants. So we need to know how to apply the norms of social consistency and social means-end coherence in these contexts. The initial answer is that if there is sufficiently deep divergence it will be unclear how to apply these social norms. A group engaged in shared deliberation concerning national economic policy, to take an example from current American politics, may well be stymied by fundamental differences in relevant empirical belief. But in such cases the participants can sometimes solve this problem for their sociality by arriving at shared policies of acceptance. If there are relevant shared policies of acceptance, they will supplement -- and, if there is divergence, displace -- the corresponding beliefs of the individuals, for purposes of assessing social consistency and social coherence.

Our ability to arrive at such shared policies of acceptance enhances our ability to reason and to act together in contexts of divergence in relevant individual belief and judgment. Such shared policies of acceptance provide part of the background with respect to which relevant forms of social consistency and coherence are assessed. And such shared policies of acceptance provide elements of the framework within which shared deliberation can, if all goes well, successfully proceed.

7. Shared policies of social rationality

Recall that the emergence of modest sociality involves, in part, the emergence of an explanatory role of norms of social rationality of intention: norms of social consistency, social agglomeration, social coherence, and social stability. I have already discussed two such (as I called them) explanatory modes of social rationality. There is, first, the basic case in which each participant guides her thought and action in light of

basic norms of individual intention rationality and in which, given the special contents and inter-relations of the intentions involved in modest sociality, there is thereby rational pressure for conformity to the cited social norms. And there is, second, the case in which the individual participants themselves proceed to internalize these norms of social rationality.

We are now in a position to describe yet a third explanatory mode of social rationality: the participants might go on to share a policy of conformity to these social norms of intention rationality. The step from the second to this third explanatory mode is the step from each having an individual policy (perhaps in a context of common knowledge of those policies) in favor of the group's conformity to these social norms of intention rationality, to a shared policy in favor of these norms, a shared policy of the sort we have been exploring.[282] Such shared policies are not focused on the "overall rational unity" of the group,[283] but only on relevant forms of intention rationality: this is another reflection of the priority of intention in modest sociality. Nor is this step to such shared policies itself a step to norms of mutual obligation. Nevertheless, these shared policies can serve as a basis for associated shared advice and criticism.

Given the nature of these social rationality norms it is likely that the interdependence involved in such a shared policy in their favor would primarily be located in box A: it would primarily be top-row, enabling interdependence. But this would still be a genuine form of sharing: it would be true that we share a commitment to these norms and that this commitment is poised to provide a framework for our thought and action. Returning to Hart's phrase, these norms would be for us a "common public standard". These common (for us) public standards would ground associated forms of shared advice and criticism; and this would lend further support to conformity to these social rationality norms. This would be a further rational mechanism in support of our guidance by and conformity to norms of social rationality of intention.

---

[282] The grounds for this step might include both instrumental and non-instrumental considerations in ways that to some extent parallel what we said in discussing the normative force of the basic individualistic norms in Chapter 1 section 3.

[283] The quoted phrase is from Carol Rovane, The Bounds of Agency: An Essay in Revisionary Metaphysics, esp. Chapter Four. Nor do these shared policies focus on what List and Pettit call "a single system of belief and desire". (34)

Conclusion: Interconnected Planning Agents

My central questions have been: How should we understand small-scale modest sociality? What resources – conceptual, metaphysical, and normative – should we draw on? How are social agency and associated norms of social rationality related to individual agency and associated norms of individual rationality?

I have tried to answer these and related questions by building on the planning theory of individual agency. The planning theory of individual agency highlights central roles and norms of intentions, understood as plan states. Appeal to this web of roles and norms supports the idea that intention is a distinctive attitude. With this planning theory in hand I have sought to show that these planning structures enable us to provide adequate resources – conceptual, metaphysical and normative – for an account of sufficient conditions for modest sociality. This is an aspect of the fecundity of planning structures.

I have tried to develop these ideas by way of a version of the Gricean methodology of creature construction. I have aimed at a construction that is conservative in the sense that it does not require metaphysical, conceptual, or normative elements that go beyond those available within the planning theory of individual agency.[284] And I have seen such a conservative construction as supporting a deep continuity between individual planning agency and modest sociality.

In pursuit of this conservative construction of shared intention and modest sociality, I have argued that the basic thesis provides relevant sufficient conditions. Returning in particular to the compressed version of that thesis, the claim is that the following provides sufficient conditions for our shared intention to J:

A. Intention condition: We each have interlocking and reflexive non-attenuated intentions in favor of our J-ing by way of each of these intentions and mutual responsiveness in sub-plan and action, and so by way of sub-plans that mesh.

B. Belief condition: We each believe that the intentions of each in favor of our J-ing will be jointly effective, in part by way of relevant mutual

[284] With the possible exception of an appeal to common knowledge.

responsiveness; and we each believe that there is interdependence in persistence of those intentions of each in favor of our J-ing.

C. Interdependence condition: There is interdependence in persistence of the intentions in favor of our J-ing.

D. Common knowledge condition: It is common knowledge that **A.-D**.

And, further, there is shared intentional activity, and so modest sociality, when such a shared intention to J leads to our J-ing in accordance with:

E. Mutual responsiveness condition: our shared intention to J leads to our J-ing by way of mutual responsiveness in sub-intention and action that tracks the intended joint activity, in conditions of common knowledge.

Such modest sociality involves the proper, rational functioning of planning structures of individual agents, structures that involve appropriately inter-related intentions and beliefs with appropriate contents, all in a context of common knowledge. Modest sociality emerges, both functionally and rationally, from these structures of interconnected planning agency.

The appeal here is to intentions; and intentions are distinctive. These intentions need to have appropriate contents and to be inter-personally inter-connected. Some of these inter-connections are built into the contents themselves: the intentions are semantically interlocking, and the intentions favor mutual responsiveness and so meshing sub-plans. Some of these inter-connections are a matter of relations between the intentions that are not primarily a matter of the content of these intentions: in particular, these intentions are interdependent in persistence and they work their way through to shared action by way of relevant mutual responsiveness. These various interconnections go beyond a solely cognitive inter-connection, as in common knowledge. But these interconnections do not essentially involve mutual obligations between the participants – though mutual moral obligations are common in modest sociality and fundamental in our social lives.

We understand such structures of interconnected planning agency in part by appeal to central norms of individual intention rationality: norms of consistency, coherence, agglomeration, and stability. Proper functioning of such structures involves

guidance by the at-least-implicit acceptance by each agent of these norms of individual rationality. Corresponding social norms of social consistency and coherence, social agglomeration, and social stability are anchored in the interaction of these norms of individual rationality with the distinctive intention-contents cited by the (compressed) basic thesis. The participants may also go on to internalize these social norms, and (even further) to arrive at a shared policy in their favor.

Within this model of modest sociality, shared intention plays a characteristic explanatory role: In modest sociality it is because there is such a shared intention that there is relevant, coordinated joint activity, related coordinated planning, and (in many cases) related coordinated bargaining. There may also be relevant shared deliberation if the parties manage to have relevant shared policies about weights. Such shared policies about weights are a special case of shared policies of acceptance, and these shared policies are understood in a way that extends the structure of the (compressed) basic thesis.

In contrast, the presence of associated mutual obligations need not ensure that such an explanatory structure is in place; after all, people sometimes have no intention at all to do what they recognize they have an obligation to do, and people sometimes have obligations they do not recognize. Further, relevant deception or coercion on the part of the participants will frequently (though not always) undermine the existence or explanatory role of such a shared intention; and that is why such deception or coercion frequently blocks modest sociality.[285]

Let's try to locate this model of modest sociality within the space of some of the main theoretical options that have been in the background of our discussion:

(a) The (compressed) basic thesis agrees with both Gilbert and Searle that the kind of sociality we are after should be distinguished from mere strategic interaction and equilibrium in a context of common knowledge. The (compressed) basic thesis claims – contrary to objections from Petersson, and from Gold and Sugden – that it does indeed provide sufficient conditions for modest sociality, in contrast with mere strategic interaction.

[285] As noted, even in special cases in which coercion or deception do not strictly block shared intentionality, they will block shared cooperativeness, in a sense of cooperative that is to some extent moralized.

(b) In providing such sufficient conditions, the (compressed) basic thesis emphasizes the central role of relevant intentions of each of the individual participants. Here it agrees with a similar emphasis in Searle's view.[286] And here it rejects Gilbert's "disjunction condition" according to which there can be shared intention in the absence of "correlative personal intentions of the individual parties".

(c) The intentions to which the (compressed) basic thesis refers are plan states explained by the planning theory, plan states with distinctive contents and inter-relations. This contrasts with Searle's appeal to a new and distinctive attitude of we-intending.[287] Indeed, we can see the intention condition of the compressed basic thesis as pointing to the kind of reductive sufficient conditions for a Searle-type "we-intention" that Searle insists are not available.

(d) The (compressed) basic thesis agrees with Gilbert that modest sociality essentially involves inter-personal interconnections that go beyond common knowledge. And this emphasis on relevant inter-personal interconnections goes beyond Searle's appeal to special attitudes of each of the individuals. In contrast with Gilbert, however, the (compressed) basic thesis tries to understand these interconnections in terms of resources made available by the planning theory of individual agency, and without an essential appeal to relations of mutual obligation.[288] The (compressed) basic thesis can, however, acknowledge the significance of familiar forms of moral obligation to cases of modest sociality.

(e) For Gilbert, the fundamental phenomenon is that of "joint commitment". And her view is that the joint-ness of a joint commitment is itself a primitive, one that does not admit of further reductive analysis. It is not analyzed in terms of the idea of a plural subject since, on the privileged reading of Gilbert's views, the idea of a plural subject is, at bottom, the idea of a joint commitment. And it is not analyzed in terms of relevant mutual obligations, since joint commitments are seen as

---

[286] In this respect there is also agreement with views of Raimo Tuomela. See his The Philosophy of Sociality: The Shared Point of View (Oxford University Press, 2007).

[287] There is here also a contrast with Tuomela's view in his 2007 book, given his emphasis on what he calls the "we-mode" and "we-mode mental states," together with his view that "the we-mode is not reducible to the I-mode." The Philosophy of Sociality: The Shared Point of View, at pp. 9-10.

[288] Nor is there an essential appeal to asymmetric authority relations, in contrast with Abraham Sesshu Roth, "Shared Agency and Contralateral Commitments," Philosophical Review 113 (2004): 359-410.

grounding relevant mutual obligations. In contrast, the (compressed) basic thesis aims providing at reductive sufficient conditions for the joint-ness or shared-ness at work in modest sociality, sufficient conditions that are continuous with the resources involved in the planning theory of individual agency. This reductive, multi-faceted model of joint-ness contrasts with Gilbert's non-reductive, all-or-none model.

(f) At the bottom of (b)-(e) is the way in which the (compressed) basic thesis draws on the planning theory of individual agency. The (compressed) basic thesis thereby supports the idea that there is a deep continuity – conceptual, metaphysical, and normative – between individual planning agency and modest sociality. There is this continuity even though it remains possible for there to be planning agents who do not have the capacity to participate in modest sociality. In contrast, both Gilbert and Searle, in different ways, see the step from individual to shared agency as bringing with it a fundamentally new conceptual and metaphysical (and, in Gilbert's case, normative) element, one that goes beyond common knowledge.

(g) By way of its use of the planning theory, the (compressed) basic thesis ties the norms of social rationality that are central to modest sociality to norms of individual intention rationality.[289]

(h) The (compressed) basic thesis claims to provide sufficient conditions for modest sociality without appeal to a new practical primitive such as Searle's we-intentions or Gilbert's joint commitments. If this claim is correct then there is an Ockham's-Razor presumption against supposing -- as do Gilbert and Searle -- that there is such a fundamentally new practical phenomenon.

Having located the (compressed) basic thesis within this solution space, let me now go on to highlight, in overview, some further, important features of that thesis.

The (compressed) basic thesis emphasizes the central role of intention-like attitudes in modest sociality: the participants need shared intentions, plans and

[289] Abraham Sesshu Roth expresses reservations about this strategy of deriving relevant social rationality norms from norms of individual intention rationality in his "Practical Intersubjectivity," in F. Schmitt, ed., Socializing Metaphysics: The Nature of Social Reality (Lanham, MD.: Rowman & Littlefield, 2003): 65-91.

policies, but they can frequently do without shared judgments or beliefs, including shared judgments or beliefs about the right and the good. This is the primacy of intention for modest sociality.

These shared intention-like attitudes will frequently be partial in the sense that they leave many matters open, some of which will need to be resolved as time goes by. And they will frequently be partial in the further sense that there is at most only partial agreement among the participants about the reasons for the sharing.[290] This is the pervasiveness of partiality in our sociality.

When there is such modest sociality there will be at least a weak kind of group agent of the shared action. It will be true, say, that we paint the house - where 'we' can here be taken to refer to a internally organized structure of the participants, a structure that can be seen as the causal source of the painting and to which relevant downstream effects can be causally attributed. The organization of this structure involves a relevant shared intention to paint. But given a moderate holism of subject-hood, it does not follow, and it is not generally true, that there is a group subject of this shared intention to paint the house. This does not follow, and is not generally true, even though aspects of the relevant inter-personal interconnections are quasi-Lockean.

Our shared intention to paint the house will many times frame our bargaining in the pursuit of mesh in sub-plans – for example, sub-plans about schedule, or distribution of tasks. It can also help frame relevant shared deliberation. When we engage in such shared deliberation our thinking together involves various shared commitments to treating certain considerations as having relevant weight within the relevant deliberative context. Perhaps we reason together in light of a shared commitment to give weight to environmental matters in choosing the paint, discarding the by-products of the painting, and so on. We can model this shared commitment to weights as a shared policy -- a shared general intention -- to give such weights in such shared deliberation. And we can apply the (compressed) basic thesis to such shared intentions/policies. Such shared policies about weights specify evaluative/normative premises to be accepted in relevant shared deliberation. And such shared policies

[290] As noted in Chapter 4, this last point contrasts with the approach of Raimo Tuomela in his The Philosophy of Sociality: The Shared Point of View (Oxford University Press, 2007) at 100-101.

about weights need not be accompanied by a corresponding convergence in judgment about these evaluative/normative premises on the part of the participants (though of course they may).

There can also be shared policies in favor of a common acceptance, in shared deliberation, of premises about what is possible and what would be effective. Shared policies about weights are a special case of such shared policies in favor of a common acceptance, in shared deliberation, of a certain proposition. Shared deliberation characteristically involves the proper, rational functioning of such shared policies. And such shared policies of acceptance can come to include shared policies in favor of social norms of intention rationality, thereby adding further support, by way of associated shared advice and criticism, to conformity to those norms.

Shared deliberation is itself a shared intentional activity. And shared deliberation involves intention-like commitments to relevant common acceptances, not merely common knowledge of associated acceptances by each. In both these respects such shared deliberation differs from the "team reasoning" modeled by Gold and Sugden.[291]

All this supports the idea that the move from individual planning agency to modest sociality, while both demanding and fecund, does not bring with it fundamentally new obstacles to locating our agency within the natural causal order. This is not by itself a solution to the problem of how to locate our individual agency in that natural order; but it is a kind of progress to see that modest sociality need not introduce yet a further problem. If we can see how individual planning agents can be the source of their own intentional activity and reasoning in the natural causal order, our theory of modest sociality will then help us to see in what sense we can be the source of our shared intentional activity and shared reasoning in that causal order. And a key to this is the planning theory of intention.

[291] See Chapter 5.