On Commitments and Other Uncertainty Reduction Tools

John Michael, Elisabeth Pacherie

On Commitments and Other Uncertainty Reduction Tools

John Michael (Central European University, Budapest)
&
Elisabeth Pacherie (Ecole Normale Supérieure, Institut Jean Nicod, Paris)

Penultimate Draft
Forthcoming in the Journal of Social Ontology

Abstract

In this paper, we evaluate the proposal that a central function of commitments within joint action is to reduce various kinds of uncertainty, and that this accounts for the prevalence of commitments in joint action. While this idea is prima facie attractive, we argue that it faces two serious problems. First, commitments can only reduce uncertainty if they are credible, and accounting for the credibility of commitments proves not to be straightforward. Second, there are many other ways in which uncertainty is commonly reduced within joint actions, which raises the possibility that commitments may be superfluous. Nevertheless, we argue that the existence of these alternative uncertainty reduction processes does not make commitments superfluous after all but, rather, helps to explain how commitments may contribute in various ways to uncertainty reduction.

1. Introduction

Many of the significant events making up humans' social and societal life are joint actions, as when we sing a duet, prepare a meal together, perform a ritual together or collectively organize a strike. Broadly speaking, joint action can be defined as 'any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment' (Sebanz et al. 2006: 70; cf. also Butterfill 2011). While collective behavior satisfying this broad definition of joint action is also observed in many other social species, human joint action is marked by its distinctive flexibility and versatility, and by the intentional character it often takes – i.e., it is typically an intentional cooperative activity where agents' actions are consciously subordinated to intentionally shared

goals. For this reason, many philosophers have been interested in characterizing a narrower, characteristically human, class of joint actions. A key concept that has emerged in this context is that of a 'shared intention'. Attempts to cash out the notion of shared intention are attempts to cash out what it takes for agents to act in a jointly intentional manner. However, there is much disagreement on how best to construe the notion of a shared intention. While the majority of philosophers agree that a shared intention cannot be reduced to a mere summation of individual intentions, they tend to disagree about what more is needed, and thus about what makes a shared intention shared. Some hold that what is special about shared intentions has to do with their contents and interrelations (Bratman, 1992; 2009); others hold that shared intentions involve a *sui generis* attitude[1], such as a *we*-intention (Searle 1990), an intention in the *we-mode* (Tuomela 2007) or a participatory intention (Kutz 2000); others identify the mark of shared intentionality at the level of the subject, holding that shared intentions can only be attributed to a plural subject (Gilbert, 1992, 2009; Helm, 2008); and yet others hold that shared intentions involve a specific form of practical reasoning, so-called team reasoning (Gold and Sugden 2007), or are constrained by special norms of group-rationality (List and Pettit 2011).

One important point of contention – which will be the focus of the present paper – concerns the nature and role of the commitments associated with shared intentions. Are these commitments different in kind from the commitments already present in the individual case? Are these commitments constitutive of shared intentions? What is their role or function? What is it about commitments that allows them to fulfill this role? Can this role be filled in other ways?

We will proceed as follows. Sections 2-4 set the stage. In section 2, we characterize the general structure of commitments and propose a typology of forms of commitment. In order to clarify what is at stake in the debate about the place of commitment within joint action, we then (section 3) contrast the respective positions of Michael Bratman and Margaret Gilbert. Alongside the important differences between their positions which we identify, we also observe a fundamental agreement that commitments are prevalent features of human joint action. In order to explain why that is so, we next turn our attention to the question of what function commitments perform within joint action. In order to address this question, we start (section 4) by considering the coordination demands that are specific to joint action, and

[1] i.e. a sui generis attitude at the individual level; in other words, these views leave open the possibility that an individual (even a brain in a vat) may have a we-intention or an intention in the we-mode in the absence of a second agent.

discuss three types of uncertainty (motivational, instrumental and common ground) that can jeopardize coordination. In the remaining sections of the paper (5-9), we assess the merits and limitations of the idea that a central function of commitments is to reduce uncertainty, and that this accounts for the prevalence of commitments in joint action. While this idea is prima facie attractive, it faces two problems. First, commitments can only reduce uncertainty if they are credible, and accounting for the credibility of commitments proves not to be straightforward. Second, there are many other ways to reduce uncertainty, in which case commitments may be superfluous (at least with respect to the function of uncertainty reduction). In section 5, we focus on the credibility problem and examine what resources Bratman's and Gilbert's accounts provide for addressing this problem. We then examine other alignment processes that can be used to reduce instrumental and common ground uncertainty (section 6) and motivational uncertainty (section 7). Finally, in section 8, we argue that the existence of these alternative uncertainty reduction processes does not make commitments superfluous after all but, rather, helps to explain how commitments may contribute in various ways to uncertainty reduction.

2. Anatomy of commitments

While talk of commitment is common currency in several areas of philosophy, philosophers rarely pause to analyze this notion and examine the similarities and differences between different forms of commitments.[2] While our purpose here is not to engage in a thorough discussion of commitment in general, it will be helpful to begin by characterizing the notion as it pertains to the phenomenon of joint action.

To begin with, the notion of commitment that both Michael Bratman and Margaret Gilbert are concerned with is *volitional*.[3] A commitment is, in Gilbert's phrase (Gilbert, 2006), a 'creature of the will' in two important respects. First, a commitment is *by the will* in the sense that an exercise of the will produces it. Second, a commitment is *of the will* in the

[2] For a recent exception, see Shpall (2014).

[3] The term 'volitional commitment' is borrowed from Shpall (2014), Cf. also Bratman (1987). Volitional commitments can be contrasted with what one might call cognitive commitments, where the latter are centrally concerned with beliefs and the norms of rationality that relations among beliefs are subject to. Thus, when we say that in believing that *p*, we are committed to believing what follows from *p*, we are talking about cognitive rather than volitional commitments. The distinction we draw between volitional and cognitive commitments overlaps partially but does not coincide with Shpall's distinction between moral and rational commitments.

sense that it binds the will in a certain way. A commitment also has a content, and the fact that it is a creature of the will imposes constraints on what this content can be: one can only commit to things that fall within the purview of one's voluntary control. This means, on the one hand, that one can form a commitment to, say, abstain from food for the next 24 hours, but that one cannot sensibly form a commitment to not feel hungry after 24 hours without food. On the other hand, it also means that one can only commit to someone else's doing something to the extent that one holds control or authority over that person's behavior. Thus, parents can commit to their children attending school regularly because they have authority over them, but they can't commit to their neighbors' children attending school regularly. Finally, a commitment to *F* is idle and unnecessary if one expects F to obtain whether or not one commits to. Thus, I need not form a commitment to brush my teeth before going to bed if I expect my teeth-brushing habits are so well entrenched that I will brush my teeth in any case.

In addition to having authors, commitments also have recipients, i.e. the person or persons you commit to. One way to start constructing a typology of commitments is in terms of whom their *authors* or *recipients* are. Herbert Clark (2006) proposes such a typology, and here we use a slightly modified and expanded version of his typology. First, the author and the recipient of the commitment can be one and the same person (*self-commitment*), or they can be different people (*other-commitments*). Second, self- and other-commitments can be either *private* or *public*. A private commitment is a commitment known only to its author and recipient. In the case of private self-commitments, since the author is the same person as the recipient, this means that only one person knows of the commitment. A public commitment, in contrast, is a commitment that has an *audience*. For instance, Peter may commit to exercising more but tell no one (private self-commitment) or he may tell his family (the audience) about his commitment (public self-commitment). The key difference between audience and recipient is that a commitment creates obligations (and corresponding entitlements) vis-à-vis the recipient but not vis-à-vis the audience. Telling his family about his commitment to exercise more (public self-commitment), may make failure to do so more costly for Peter, who would incur the risk of teasing and public embarrassment, but his family, as a mere audience to the commitment, would not be entitled to his delivering on it. In contrast, if Peter was to commit to his family to exercise more (e.g. make a promise to them), he would have an obligation to them, and they could hold him responsible for his failing to fulfill his commitment.

The examples we have considered so far are examples of *unilateral* commitments.

However, both self- and other-commitments can also be *interdependent*. Suppose there is strong sibling rivalry between Peter and his sister Sarah. Each is strongly motivated to outcompete his or her sibling in every domain. Thus, while neither is particularly fond of school, Peter commits to working hard at getting good grades just so long as Sarah so commits as well, and vice-versa, and they make no secret of it. This is a case of interdependent public self-commitments. If instead, Peter and Sarah had agreed that he would commit to her to work hard at getting better grades in math just so long as she would commit to him to work hard at getting better grades in history and vice-versa, the proposed commitments would be interdependent other-commitments. If Peter doesn't work hard at improving his math grades, Sarah can hold him responsible for his own failure and consider herself released from her commitment to him, and vice-versa.

Finally, within the category of interdependent commitments, we may want to distinguish between commitments that are merely *bilateral* and commitments that are *joint*, where the latter differ from the former in that they involve a shared goal. The examples of interdependent commitments we have considered so far did not involve a shared goal, and were thus simply bilateral (and not joint commitments). In the rival siblings example, Peter's goal is to outcompete Sarah and Sarah's goal to outcompete Peter. These goals are incompatible and clearly cannot be shared. Similarly, in the second example Peter's goal is to improve his math grades, and Sarah's to improve her history grades. While their goals are compatible, they are not shared. Rather, they constitute two distinct outcomes, each of which could be brought about independently of the other. But suppose now that Peter and Sarah actually both want their parents to be pleased with them, and know that Peter's math grades and Sarah's history grades have to improve in order to achieve this. Their commitments to improve their grades are now not just conditional on each other but also conditional on both being committed to pleasing their parents. In this case, they take on interdependent other-commitments to a shared goal. In other words, they engage in a joint commitment.

With this rough typology now in place, let us now consider theoretical options for conceptualizing the role of different kinds of commitment in joint action. As we move forward, it will be especially important to distinguish between substantive differences among leading theoretical accounts, on the one hand, and merely terminological differences, on the other. If nothing else, our rough typology will be helpful in this regard.

2. Bratman vs. Gilbert on commitments in joint action

Margaret Gilbert 's chief argument in favor of her claim that shared intentions essentially involve joint commitments may be called the argument from rights and obligations. Gilbert (2006; 2009) bases her argument on observations concerning the way people talk and think about shared intentions in everyday settings. From these observations she derives three criteria of adequacy for accounts of shared intentions. First, she notes that agents that are parties to a joint activity understand that they have a standing to demand of one another that they act in a manner appropriate to their joint activity, and to rebuke the other party should they act in a manner inappropriate to it. They understand, moreover, that this standing has its source in the joint activity itself. Gilbert considers this as evidence that obligations and rights are part and parcel of joint activity as such; that is, participants in a joint activity have obligations towards each other to act in conformity with their shared intentions and correlative entitlements or rights to others so acting. Thus, Gilbert takes it to be a criterion of adequacy for an account of joint action and shared intention, that it explain the presence of obligations in joint actions and explain how these obligations are grounded in the joint activity itself.

In addition to this obligation criterion, she proposes a second criterion of adequacy, which she also takes to derive from our intuitive understanding of what is involved in acting jointly. Barring special provisions to that effect, no one party to a joint action is in a position to unilaterally decide on the details of the joint action, make changes to a joint action plan or break off from the joint activity. Rather, the concurrence of all parties is required for any such modification. Thus, according to the concurrence criterion, ‘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, or that a given party be released from participating in it’ (Gilbert, 2009, 173).

Finally, Gilbert (2009) also proposes a third criterion of adequacy, the disjunction criterion, according to which ‘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’ (p. 172). Gilbert has in mind cases such as this: you and I formed a shared intention to walk together to the top of a hill. Half-way through, both of us feel tired and no longer have any intention of hiking to the top of the hill, but neither of us has as yet said anything about their change of mind to the other. According to Gilbert, it is intuitively true that at that point we still have a shared intention to walk together to the top of the hill, despite our now lacking individual intentions to do so. In our view, this disjunction criterion can be seen as a special case of the concurrence criterion. The

concurrence of all parties is required in order for a given shared intention to be rescinded, but of course a precondition of concurrence in the sense intended by Gilbert is that all parties know where all others stand. In the example given, this pre-condition does not obtain.

Gilbert then argues that – in contrast to an account of shared intention that appeals to a structure of correlative individual intentions, such as Bratman's – an account of shared intentions that is based on the notion of a joint commitment satisfies these criteria of adequacy. According to her joint commitment account:

> Persons X, Y, and whatever particular others share an intention to do A if and only if X, Y, and these particular others are jointly committed to intend as a body to do A. (Gilbert 2009: 179)

Gilbert describes the idea of a joint commitment as an analogue of the idea of personal commitment in individual agency. When an individual has formed an intention or made a decision, he has in virtue of this intention or decision sufficient reason to act in a certain way; that is, all else being equal, he is rationally required to act in that way. Thus, a personal intention or decision entails a personal commitment to act in a certain way. Analogously, a joint decision or intention to act involves a joint commitment. Importantly, however, Gilbert insists that joint commitments are not concatenations of personal commitments. Rather, in the basic case, a joint commitment is created when each of two or more people openly expresses his personal readiness jointly with the other to commit them all in a certain way, and it is common knowledge between them that all have expressed their readiness. The author of a joint commitment comprises those who have jointly committed themselves by their concordant expressions. Together they constitute the plural subject of the commitment. In Gilbert's view, plural subjects and joint commitments are constitutively linked: there can be no plural subjects without joint commitments and there can be no joint commitments that are not the commitments of a plural subject.

Gilbert (2009) glosses what she means by ‘intending as a body to do A’ as emulating, as far as possible, by virtue of the actions of each, a single body (or agent) that intends to do the thing in question. One way to flesh out the idea of ‘intending as a body’ is in terms of satisfying the types of rationality constraints that bear on individual agency. To ‘intend as a body’ would then be a matter of intending to act in such a way that the actions of each together satisfy norms of consistency, agglomeration and means-end coherence. Speaking of ‘intending as a body’ also conveys the idea that a party to a shared intention may intend to do A *qua* member of that body while possibly lacking a personal intention to do A, i.e., an

intention to do A *qua* individual.

Once we take joint commitments to constitute the core of shared intentions, it is easy to understand how shared intentions can meet her three criteria. Commitments in general have normative force: they create obligations for their authors and concomitant rights for their recipients. The author of a commitment can be relinquished from his obligation to act in conformity with the commitment only if the recipient of the commitment waves their right to conforming action. If, as Gilbert claims, shared intentions involve not just commitments but joint commitments, and have plural subjects as both authors and recipients, then in forming a joint commitment, the parties to the commitment together impose obligations on each other to act in conformity with the commitment, and concomitant rights to demand of one another that they so act. The obligation criterion is thus fulfilled. In addition, since a commitment can only be rescinded with the consent of its recipient and the recipient is the plural subject, the joint commitment account of shared intention also satisfies the concurrence criterion. This entails that a shared intention may persist even when individual parties to it no longer have correlative personal intentions, so the disjunction criterion is also fulfilled.

Importantly, Gilbert also argues that the obligations created by joint commitments for those jointly committed can neither be construed as forms of moral obligations, nor be derived from the constraints inherent in individual rational planning. Their normativity is therefore neither that of moral norms nor that of norms of individual rationality. It is, rather, a *sui generis* form of social normativity. The notion of joint commitment is thus basic or non-reducible insofar as joint commitments are the source of this *sui generis* form of social normativity.

Michael Bratman (2009) agrees that mutual obligations and entitlements are very common in joint action. However, in contrast to Gilbert, he thinks that these obligations are typically the familiar kinds of moral obligations associated with assurance, reliance, and promises, and he's not convinced that they are essential to joint action and shared intention. Instead, Bratman develops a constructivist approach to shared intentions that exploits the conceptual and normative resources of his planning theory of individual agency. His aim is twofold: to show that the kind of normativity central to shared agency can be derived from the normativity already present in individual planning agency, and to show that one can capture the interconnections among agents characteristic of shared agency by construing shared intentions as complexes of interlocking intentions of individual agents.

Bratman's planning theory of agency (Bratman, 1987) stresses the commitment to action that is a distinctive characteristic of intentions. Individual intentions are commitments

to act. These commitments have both a volitional and a reasoning-centered dimension. Their volitional dimension concerns the relation of intention to action: intentions are conduct-controlling pro-attitudes, whereas ordinary desires are merely potential influencers of action. Their reasoning-centered dimension concerns the norms of practical rationality they are subject to. First, once we have formed an intention to do *A*, we see the question whether to do *A* as settled. In the absence of relevant new information, an intention is rationally required to resist reconsideration (*stability*). Second, when we form an intention, our plans are typically only partial to begin with, but if they are to eventuate in action, they need to be filled in prior to their execution. Thus, in intending to do *A* we are also committed to reasoning about means, preliminary steps or more specific courses of action. This intention specification process is rationally required to meet *consistency constraints*: our plans should be internally consistent, means-end coherent and consistent with our beliefs about the world. Finally, the volitional and reasoning-centered dimensions of intentions together account for another important function of prospective intentions, namely to support both *intrapersonal and interpersonal coordination*. Because intentions have stability, are conduct-controlling, and prompt reasoning about means, they support the expectation that we will do tomorrow what we intend today to do tomorrow, and thereby also impose constraints on what other intentions we can have. Our intentions are thus also rationally required to be agglomerative: it is rational to intend to *A* and to intend to *B* only if it is rational to intend to *A* and *B*. This would involve, in Bratman's terms, commitments to mutual compatibility of relevant sub-plans, commitments to mutual support, and joint-action tracking mutual responsiveness.

Bratman's key idea is that we can capture the forms of normativity and the connections among participants distinctive of shared agency if we construe shared intentions as complexes of interlocking and interdependent intentions of individual participants, where these intentions interlock in the sense that each intends that the joint activity go in part by way of the relevant intentions of each of the other participants. These intentions of individual participants, in responding to the norms of practical rationality governing individual planning agency, will normally support the norms of social agglomeration and consistency, social coherence and social stability to which shared intentions are subject.

To recap, Michael Bratman and Margaret Gilbert both agree that there is a normative dimension to joint action, and that having a shared intention involves being committed in certain ways. Yet, they disagree about the nature of the normativity and of the commitments that are at stake. According to Gilbert, shared intentions constitutively involve joint commitments. Thus, the form of commitment they entail is, in the terms laid out in our

typology in section 2, a kind of interdependent other-commitment with a shared goal. Moreover, on her view, their normative import pertains to a *sui generis*, irreducible kind of social normativity. In contrast, according to Bratman, the only commitments that are constitutively involved in shared intentions are commitments to norms of practical rationality, and the norms of practical rationality that govern intentional joint actions are simply extensions of the norms of practical rationality already governing individual planning agency. In terms of our typology, the commitments that are constitutive of Bratmanian shared intentions are *interdependent public self-commitments*.

On the other hand, Bratman does acknowledge that joint commitments (interdependent other-commitments with shared goals) are characteristic of joint action; he just does not take them to be *necessary* for shared intentions or joint action. As he puts it: 'I believe that the normal etiology of a shared intention does bring with it relevant obligations and entitlements when the shared activity is itself permissible. But I also believe that this etiology is not essential to shared intention itself' (Bratman 1999, 132).[4]

In sum, the divergent views espoused by Bratman and Gilbert belie a fundamental agreement that joint commitments are (at least) a prevalent feature of human joint action. If this is so, then this fact calls out for an explanation. In seeking such an explanation, our starting point will be the conjecture that if joint commitments are prevalent features of human joint action, then it is likely that they perform some function which accounts for that prevalence. Our aim in the next section will be to home in on what that function is.

4. Coordination in joint action and uncertainty reduction: types of uncertainty

Successful joint action depends on the efficient coordination of participant agents' goals, intentions, plans, and actions. As one of us argued elsewhere (XXX, 2012), it is not enough that agents control their own actions, i.e., correctly predict their effects, monitor their execution and make adjustments if needed. They must also coordinate their actions with those of their co-agents so as to achieve their joint goal. For that they must monitor their partner's intentions and actions, predict their expected consequences and use these predictions to adjust what they are doing to what their partners are doing. The implication of these processes, however, is not unique to joint actions, nor is it enough to promote their success. In

[4] Note also that Tuomela's distinction between two forms of joint actions, I-mode joint actions and we-mode joint actions, can be seen as a way of reconciling the respective insights of Bratman and Gilbert (Tuomela, 2007)

competitive contexts they also play an important role. In a fight, for instance, being able to anticipate your opponent's moves and to act accordingly is also crucial. What is furthermore required in the case of joint action is that co-agents share a goal and understand the combined impact of their respective intentions and actions on their joint goal and adjust them accordingly. In competitive contexts, an agent should typically aim at predicting his opponents' moves, while at the same time endeavoring to make his own moves unpredictable – or, better yet, positively misleading – to his opponent. In contrast, in cooperative contexts mutual predictability must be achieved in order for efficient coordination towards a shared goal to be possible. Agents should be able to align their representations of what they themselves and their partners are doing, and of how these actions together contribute to the shared goal.

Various forms of uncertainty can undermine mutual predictability, the alignment of representations and hence coordination. They can be organized into three broad categories[5]. The first category involves *motivational uncertainty*: we can be unsure how convergent a potential partner's interests are with our own interests and thus unsure whether there are goals we share and can promote together. Additionally, even if we know what their current preferences are and that they match ours, we can be unsure how stable these preferences are. Thus, you're unlikely to undertake a walking tour of Scotland with someone you're afraid would change their mind and give up after a few days, leaving you alone in the middle of Rannoch Moor.

The second category involves *instrumental uncertainty*: even assuming that we share a goal, we may be unsure what we should do to achieve that goal, or, if we have a plan, we may be unsure how roles should be distributed among us, or, even if the plan and the distribution of roles are settled, we may be unsure when and where we should act. Thus, even if you know that Andrew would be quite willing to undertake a walking tour of Scotland with you and is an experienced hiker, and that you can count on him not to give up, you may be unsure how to proceed, what is the best time of the year for such an undertaking, what the appropriate itinerary is, how many miles a day one can expect to cover on moorlands or going up rugged mountains, who is to buy train tickets or book accommodation, and so on.

The third category involves *common ground uncertainty*: we can be unsure how much of what is relevant to our deciding on a joint goal, planning for that goal and executing our plan is common ground, or mutually manifest to us. In other words, it is not enough to ensure

[5] To be clear, this is merely a rough-and-ready distinction that serves to structure the discussion; the three proposed categories are likely to overlap significantly.

coordination that we are actually motivated to pursue the same goals and have sufficiently similar instrumental beliefs and plans regarding how these goals should be achieved. If either you or I don't think this is the case, we won't engage in a joint action. Nor is it enough that we happen to be aware that the other has the same motivations and plans, because we still may be unsure whether the other shares this awareness. Rather it must be transparent to us that our motivations and plans are aligned.

Note that to say that various forms of uncertainty can jeopardize coordination and undermine joint action is not to say that absolute certainty and complete common ground regarding relevant matters are required to engage in joint action. Human beings are fallible, and we can never be completely sure of what others actually know, understand, believe or want. In other words, uncertainty can never be completely eliminated; we can only hope to reduce it to acceptable levels, that is, to levels where the probability of misalignment is low enough that that the expected benefits of acting together outweigh the costs associated with failed coordination. So what joint action realistically requires are uncertainty reduction tools rather than uncertainty-elimination tools. The questions to be considered in the next sections are thus: how do commitments fare as uncertainty reduction tools? How do they fare compared to other uncertainty reduction tools?

5. Commitments as uncertainty reduction tools

One consequence of commitments, if they are credible, is that they make agents' actions predictable in the face of fluctuations in their desires and interests. As a result, they may enable agents to have more reliable expectations about each others' actions than would otherwise be possible. In other words, commitments may contribute to the reduction of the three types of uncertainty set out in section 4. First of all, they may reduce motivational uncertainty because they make agents willing to perform actions that they otherwise would not perform – to work, for example, given that somebody has made a commitment to pay them for it, or to lift one end of a heavy table that cannot be moved by one agent alone. More specifically, having reliable expectations about others' actions facilitates cooperation[6] and

[6] A simple definition that captures the relevant sense of cooperation here is that is a process where a group of individuals act together for their common/mutual benefit, as opposed to acting in competition for selfish benefit.

coordination[7]. In cooperation problems, such as the prisoners' dilemma game (Axelrod 1984), individuals are tempted to defect in order to maximize their own benefits but by cooperating maximize the overall group benefit. In coordination problems, on the other hand, such as the stag hunt game (Skyrms 2004), two agents each maximize their individual benefits if they coordinate their actions but get no benefit if they try to coordinate but fail to do so, and are therefore tempted to opt for a smaller benefit which does not depend upon coordinating with the other agent. Credible commitments may also reduce instrumental uncertainty because if one agent has reliable expectations not only about another agent's willingness to contribute to a goal but about the specific contribution she intends to make, it is much easier to plan a complementary contribution to that goal. In other words, reliable expectations facilitate the planning of sophisticated actions with complementary subplans, which depend upon and build upon each other, as well as the online coordination of joint actions among multiple agents. Finally, commitments can be made explicit, and may thereby reduce uncertainty about common ground. When two agents' jointly commit to pursuing a common goal according to an agreed strategy, their willingness to pursue the goal, and their intentions to do so according to a particular plan, become mutually manifest.

Thus, it is initially very plausible to think that a central function of commitments in joint action is to reduce uncertainty. However, this conjecture faces two immediate problems. First, commitments can only reduce uncertainty if they are credible, and accounting for the credibility of commitments proves not to be a simple matter, as we shall see in a moment. Secondly, there may be many other ways of reducing uncertainty, in which case commitments (even if they are credible and can therefore reduce uncertainty) may be superfluous. In the remainder of this section, we will consider the first of these problems. In the following two sections, we will examine the second problem.

First, then: why is the credibility of commitments not a straightforward matter? The problem is that it appears to be irrational to engage in and follow through on joint commitments, since they foreclose options which may arise and which may be more attractive than the action to which one is committed – in other words, commitments foreclose options which may maximize an agent's interests. Thus, if an agent makes a commitment to perform a particular action, and her interests or desires subsequently change, it is not clear why she should remain motivated to fulfill the commitment. As a result, it is not clear why anyone else

[7] The relevant sense of coordination in the context of joint action is for two or more agents to select and perform actions that complement each other in bringing about a shared goal or compatible but distinct individual goals.

should expect her to. If, for example, Sally makes a commitment to Frank, which Frank does not think Sally is motivated to fulfill, then it is difficult to see why Frank should consider the commitment to be credible and should expect Sally to perform the action they are committed to. And if Frank cannot rely on Sally's commitment, then the commitment will not be performing its function of stabilizing expectations and making them more reliable.

The problem is illustrated nicely by an example from Hume (1740, pp. 520–521): two farmers each expect a good corn crop this year. Each will need their neighbor's help in harvesting the corn, and since their crops are expected to ripen at different times, they can maximize their harvests if each helps the other when their respective crop ripens. The problem is that the farmer whose corn ripens later (call him Frank) does not help the other farmer (call her Sally) because he reflects that if he were to help Sally, then Sally would no longer have anything to gain from helping Frank later on, and would thus not help him when his corn ripens. So, since Frank does not expect Sally to aid her when the time comes, he does not help with her harvest, and of course she does not help him when his corn crop ripens either, and both are worse off than they would be if they helped each other.

In some cases, this type of problem can be solved by *externalizing* commitments. For example, ,the farmers Frank and Sally might sign a contract that entails a daunting fine for reneging on the commitment to help with the corn harvest. In this case, they change the payoff structure for the available action options, making reneging a less attractive option than it otherwise would be. This motivates each of them to stick to the planned course of action, and also makes it credible to each of them that the other will also do so. Thus, it is clear enough how commitments are motivated and therefore also credible when they are externalized, but what about cases where they are not externalized? We do not usually sign contracts when agreeing to take a walk together. Yet people often engage in and follow through on such commitments. Why do they do so?

What resources do Bratman's and Gilbert's accounts proffer for addressing this problem? Consider Bratman first. On his account, commitments to act are subject to norms of practical rationality (stability, means-end coherence, consistency, agglomeration). Hence, expectations about an agent's actions will be constrained by these norms, and their predictability will thus be enhanced. However, it is important to note that the application of these norms is premised on the assumption that an agent is rational. The problem, then, is that there is no guarantee that the agent will not modify her intentions if her interests change. Indeed, practical rationality may *demand* that she re-consider her intentions if new reasons come to light. And since it can be rational for an agent to be open to re-considering her

intentions, fluctuations in her interests (or even the suspicion thereof on the part of the other agent) threaten to unravel the network of interdependent intentions which, on Bratman's account, constitute shared intentions. To be fair to Bratman, on his account whether the nonreflective (non)reconsideration of an intention is rational for an agent depends on whether it is a manifestation of habits of reconsideration that are themselves reasonable (Bratman 1987: 64 sq.). The account allows that neglecting to reconsider an intention can be rational when manifesting reasonable habits of (non)consideration but irrational when manifesting unreasonable habits and thus endorses neither pusillanimity nor foolish stubbornness as rational. However, an appeal to reasonable habits can only explain how commitments reduce uncertainty arising from the possibility of minor fluctuations in external circumstances. And it does not afford any possibility of explaining why the social nature of some commitments (i.e. their having been made public or their being instances of one of the types of other-commitments discussed above) may reduce uncertainty.

On Gilbert's account, joint commitments are credible for a different reason, namely because joint commitments give rise to obligations, and people tend to act as their obligations dictate. Thus, on this account, the reduction of uncertainty achieved by commitments is premised on the social normative assumption that people act as their obligations dictate. This, according to Gilbert, enables shared intentions on the plural subject account (in comparison to Bratman's account) to provide a 'more stable framework for bargaining and negotiation and, relatedly, a more felicitous means of coordinating the personal intentions of individuals, and keeping them on the track of the shared intention.' (2009: 185). However, one might justifiably demand an explanation of what motivates people to act as they are obligated to, and Gilbert's account merely asserts that people are so motivated without providing an explanation of why this is so. Thus, while Gilbert's account, in contrast to Bratman's, opens up the possibility that commitments stabilize agents' behavior in the face of fluctuations in their interests, this possibility depends upon agents being motivated to fulfill their obligations. In the absence of an explanation of why and/or under what conditions agents are so motivated, Gilbert's account must be regarded as incomplete.

6. Reducing instrumental and common ground uncertainty

In the previous section, we argued that a commitment can only reduce uncertainty if it is credible, and that its credibility depends upon the motivation of the committed agent to honor her commitment. We also diagnosed some difficulties that Bratman and Gilbert have in

addressing this motivational problem. The upshot was that unless the motivational problem can be dealt with adequately, it is doubtful whether commitments can really serve the function of reducing uncertainty in joint actions. Let us now turn to the other challenge that must be answered if commitments are to be accorded a central role in reducing uncertainty, namely that many other processes appear to fulfill this function already, and that commitments may therefore be superfluous. If so, then their usefulness as uncertainty reduction tools will hardly account for their prevalence. We will take up this issue in two separate steps: in this section, we will examine alignment processes which reduce instrumental and common ground uncertainty, and in the next section we will turn to processes which reduce motivational uncertainty and which appear not to require commitments in order to do so.

There has been a great deal of work in recent years, both conceptually and empirically, investigating the cognitive processes by which shared representations are generated, and uncertainty thereby reduced, in joint action. In order to provide a sample of this literature, let us begin with a rough distinction among three classes of process that contribute to the alignment of representations in joint action contexts, and thereby to the reduction of instrumental and common ground uncertainty: automatic alignment processes, intentional alignment processes, and pre-aligned representations.

Automatic alignment processes can induce the formation of shared representations in several individuals through physical coupling or perception-action coupling, and do so independently of whether these agents intend to act jointly. Interpersonal entrainment mechanisms are one illustrative example. Thus, for example, people sitting in adjacent rocking chairs will tend to synchronize their rocking behavior, even if the chairs have been manipulated such that their natural rocking tempos are different (Richardson et al. 2007). It is likely that such processes make partners in a joint action more similar and thus more easily predictable for each other, and thereby reduce uncertainty about goals, action plans and common ground. And indeed there is some evidence that synchronization may lead to an increase in cooperative behavior in economic games (Valdesolo et al 2010). This may be due in part to a heightening of the sense of belonging to a group, which increases the perceived common ground. It is also likely that synchronization can contribute to generating rapport among participants to a joint action, and thereby to the reduction of motivational uncertainty, but let us continue to focus for the moment on instrumental and common ground uncertainty.

While interpersonal entrainment mechanisms pertain most directly to the alignment of bodily movements, there is also evidence of alignment at the cognitive level, i.e. at the level of representations that are relevant to joint actions. For example, people tend to ‘co-represent’

tasks that other people are performing next to them, even when it is counterproductive for them in that it interferes with performance of their own task (Atmaca et al., 2008; Sebanz et al., 2005; Tsai et al, 2008). Moreover, it has been shown that people tend to predict the sensory consequences not only of their own but also of other participants' actions (Wilson and Knoblich 2005), to automatically track others' (true and false) beliefs (Kovacs et al 2010; Van der Wel et al. 2014), and to automatically monitor their own and other's errors – and indeed there is an overlap in the neural areas that are activated for detection of one's own and others' errors (van Schie et al. 2004; de Bruijn et al 2009). Indeed, various other bodily and affective processes also serve to generate and sustain common ground in social interactions. Think, for example, of gaze-following, by which people automatically keep track of what other people are attending to and acquire valuable information about what actions they are considering or intending (Tomasello et al. 2004). All of these processes are important insofar as they contribute to the reduction of uncertainty about how individuals represent situations and tasks (common ground uncertainty) and about their ongoing and upcoming actions within joint action contexts (instrumental uncertainty).

Another type of automatic process that may contribute to the reduction of instrumental and common ground uncertainty can be captured with the term 'coordination smoother', i.e. any kind of modulation of one's movements that 'reliably has the effect of simplifying coordination' (Vesper et al. 2010:2). For example, one may exaggerate one's movements or reduce variability of one's movements to make them easier for the other participant to interpret (Pezzulo et al. 2011). Although coordination smoothers may in some cases be produced automatically, the term may also be applied to processes, such as nods, winks and gestures, which are produced intentionally. And of course, there are a myriad other ways in which intentional alignment processes can reduce uncertainty, with linguistic communication being the paradigmatic case.

Finally, pre-aligned representations, such as social coordination conventions, pre-established scripts and routines, institutional settings and shared task representations acquired through previous practice, may also serve to minimize uncertainty in joint actions. In addition, unless a situation is completely new to them, agents will already have a stock of relevant pre-aligned representations they can rely on to facilitate coordination. The effect of pre-aligned representations is that there is less uncertainty to begin with, and thus also less need to reduce it.

Some of these alignment processes (e.g., signaling and communicative actions) are akin to commitments. Yet, they are typically unilateral commitments rather than joint

commitments. Others, however, are not commitments (or commitment-like) in either Bratman's or Gilbert's sense. First of all, they are involuntary processes and thus not 'creatures of the will'. Secondly, they are sub- or a-rational psychological processes, and thus outside the scope of Bratman's account. And thirdly, they have no deontic dimension.

These alignment processes may play a crucial role in attempts to spell out what exactly 'acting as a body' involves and how it is possible. They can also be seen as important tools for reducing instrumental uncertainty and for some (e.g., joint attention) common ground uncertainty. But can alignment processes that do not rely on commitments also contribute to the reduction of motivational uncertainty? Perhaps, as we shall see in the next section.

7. Reducing motivational uncertainty

While there has been a great deal of work on various types of processes which align agents' representations and thereby reduce instrumental and common ground uncertainty within joint actions, much less attention has been paid within the literature on joint action to processes which reduce motivational uncertainty. However, we do not have to look too far afield in order to find relevant empirical findings and theoretical proposals that bear upon this issue.

Behavioral economics, for example, has produced a wealth of relevant research. One issue that continues to command great interest in this domain is the need to explain the existence of human cooperation, as evinced, for example, by people's tendency to behave far more generously than their narrow self-interests dictate in various economic games, such as the prisoners' dilemma and the dictator game (Camerer, 2003; Rand and Nowak, 2013). In order to account for this, we need a two-tiered story: first of all, we need a mechanism for the evolution of cooperation, i.e. 'an interaction structure that can cause cooperation to be favored over defection' (Rand and Nowak, 2013). Several such mechanisms (direct reciprocity, indirect reciprocity, spatial selection, multilevel selection, kin selection) have been proposed and put to test in laboratory experiments and field studies. Secondly, we need an account of the psychological processes that implement these cooperative strategies. Let us briefly consider two types of psychological processes that may play this role: group identification and moral emotions.

On Bacharach's (2006) 'team-reasoning' account, group identification can lead individuals to be motivated to act in ways that optimize the interests of the group with which

they identify, even at the expense of their own individual interests.[8] For Bacharach, whether or not an agent identifies as a member of a group or team is a matter of what frame she uses to represent herself and the agents with whom she is interacting. What frame she uses, in turn, is not a matter of choice but of involuntary psychological processes. The adoption of a we-frame involves an agency transformation: the person who self-identifies as a member of a team thinks of his or her agential self as a component part of the team. This can lead her to assess action options according to the benefits they bring to the team of which she is a part, and to construct plans that aim to bring about the greatest possible benefit for the team. In other words, the adoption of a we-frame primes *team-reasoning*, which yields cooperative behavior. For this reason, Bacharach takes group identification to be a basic human propensity. As he puts it, group identification may be 'the key proximate mechanism in sustaining cooperative behavior in man' (2006: 111).

When considering Bacharach's approach, it is natural to pose the question of how to identify conditions for the production of group identification. Bacharach appeals to research on group identification in social psychology (Brewer, 2003; Tajfel, 1981; Turner *et al*., 1987). that has identified a number of conditions that tend to produce group identification, including: belonging to the same social category (e.g. being a woman, a philosopher, a Parisian); becoming the same *ad hoc* category (being born on the 1st of June), face-to-face contact, 'we' language, shared experience (e.g., being an air crash survivor), having common interests, being subject to a common fate, interdependence, and a competing outside group (e.g. analytic vs. continental philosophers). Whether a situation promotes group identification – and if so, to which group – depends on whether the situation presents some of these properties, and whether they are salient enough to induce the corresponding group frame. The minimal group methodology, originally developed by Taifel and colleagues (Taifel, 1970; Taifel et al., 1971; Billig & Taifel, 1973) and widely used in social psychology, reveals how easy it is to create groups and induce group identification. Randomly assigning individuals to one of two groups ostensibly on the basis of arbitrary and virtually meaningless distinctions (e.g., overestimators vs. underestimators of the number of dots in a display; preference for certain abstract paintings over others, or even toss of a coin) was enough to induce group identification and in-group favoritism.

[8] For discussions of how Bacharach' approach can yield an account of shared intentions altogether different from either Bratman's or Gilbert's, see Gold & Sugden (2007) and Pacherie (2011, 2013).

Note, however, that to get an account with real explanatory leverage, we also need a motivational story. Why are we motivated to group-identify in the first place? This question leads us to a second class of psychological processes which may serve as proximate mechanisms supporting cooperation: namely social emotions, such as fellow feelings or sense of belonging, and moral sentiments, such as guilt, shame, sympathy, trustworthiness and outrage[9]. There are two immediate reasons why emotional processes are useful and effective tools for promoting cooperative behavior: first of all, they are automatic, involuntary processes; and secondly, they are intrinsically motivating and are the proximate cause of most behavior. Taken together, these two features make it possible for emotions to influence an agent's behavior independently of her conscious assessment of what is in her best interests. In other words, they provide the insulation from narrow self-interests which is required to support cooperative behavior.

And indeed there is evidence that once children are able to make a distinction between self and other, as evidenced by their ability to recognize themselves in a mirror, they also begin to react with empathic and sympathetic responses to victims of distress and with appropriate, other-directed comforting and prosocial behavior (Bischof-Köhler, 1991; Zahn-Waxler et al., 1992; Eisenberg and Fabes, 1998), and that sympathetic arousal is induced in 2 year-olds (as measured by pupil dilation) when others are in need of help, and subsides when help is offered, regardless of who offers the help (Hepach et al., 2012). And in adults, experiments using economic games (Rand & Nowak, 2013) provide evidence that affective, intuitive processes support cooperation in one-shot games, while reflection and deliberation lead to selfish behavior. For example, induction of an intuitive mindset through priming or time pressures increases cooperation in Public Goods Games relative to a more reflective mindset (Rand et al., 2012). Similarly, increasing the role of intuition through cognitive load augments generosity in a resource allocation game and in the Dictator game (Roch, et al., 2000; Schutz et al., 2012). Cornelissen et al (2011) replicated this finding, and also found evidence that perceived interpersonal closeness and perceived similarity also correlated with generosity – especially when participants did not reflect on their decision. It is also interesting to note that Bonnefon et al. (2013) reported evidence that trustworthiness detection is an automatic process, independent of general intelligence and effortless.

[9] For present purposes, we will adopt Jesse Prinz's (2004) conception of sentiments as emotional dispositions. Thus, for example, if I have a sentiment of moral disapprobation toward a type of action, this disposes me to experience different emotions depending on whether the action is performed (outrage), averted (happiness), threatened (fear), etc.

Now, although the foregoing remarks may help to make it plausible that affective processes increase people's motivation to engage in cooperative behavior, they have done nothing to address the issue about whether such processes reduce *uncertainty* about people's motivation. We can distinguish two (mutually compatible) options for addressing this. One option would be to argue that, contra the Hobbesian idea found in Gilbert (1989) that mutual distrust is the default expectation, the default is rather an expectation of trust and cooperation. If so, then the processes described so far would not so much reduce uncertainty as explain why it is relatively unproblematic for people to operate with a default expectation of trust and cooperation. In other words, the explanation would not be an explanation of how uncertainty is reduced but of why there is less uncertainty in the first place than one would expect on the basis of individuals' rational calculation of their expected benefits.

A second option would be to give an account of how either one's own emotions or one's attribution of emotions to others reduces uncertainty about motivation. And indeed such an account can be given, as it has been suggested that moral sentiments may play an important role not only in stabilizing cooperative behavior but also specifically in doing so by solving commitment problems (Frank 1988; Sterelny 2003). A commitment problem arises when an agent would stand to gain if she could make a credible commitment, but when she has a countervailing motivation which undermines the credibility of her commitment. An important class of commitment problems pertains to cooperation, for commitment problems arise when multiple agents would benefit by cooperating but where at least one agent has a temptation to defect, which undermines the other agent's confidence that she will cooperate. The problem, then, resides in the difficulty of making a credible commitment (to cooperate rather than to defect) when there is a countervailing motivation. It must be emphasized that commitment problems do not always pertain to cooperation. Indeed, an important class of them pertain to *threats*. To illustrate this, consider the following toy example: Barbara owns a shop that sells umbrellas, and Sam is considering whether to steal an umbrella from the shop. He considers that if he gets caught, he can simply deny it, and Barbara will not bother to pursue the matter, because it would cost her more time and energy than an umbrella is worth. So even if Barbara makes a pronouncement committing her to prosecuting all thieves, it is not sufficiently credible to deter Sam. One basic strategy for solving such problems is to change the payoff structure of one's options. If, for example, one signs a contract committing one to cooperate, which entails a fine for defecting, then defecting no longer has a higher payoff than cooperating, and the commitment to cooperate is credible. Similarly, Barbara may join a club of shop owners and sign a contract binding her to prosecute all thieves, and otherwise to pay a

large fine. The large fine that she would risk having to pay if she did not prosecute Sam now alters the payoff structure of her options, making it far more credible that she would actually prosecute him. The important thing to emphasize here is that if the commitment is credible, then Sam will not test her in the first place, and she will not have to spend the time and energy prosecuting him or pay the fine.

Now, what do emotions have to do with this? The general idea is that the anticipation of emotional outcomes of actions changes the payoff structure for an agent's action options in the same way as a contract to cooperate or Barbara's contract with the other shop owners. For example, if one agent could get away with cheating (i.e. she could be confident that she could avoid detection), she may nevertheless refrain from doing so because she wants to avoid the negative emotional outcome that she expects to ensue from cheating (e.g. guilt, shame). If other potential collaborators take her to be a person given to such emotions, then they may also prefer to interact with her because of this. This may be either because of past experiences with her, because of her reputation, or because of some aspect of her appearance or behavior. One may even speculate that this is part of the reason why emotions are often associated to observable physical symptoms that are costly or difficult to fake (e.g. blushing). The same goes for threats: if Barbara is believed to be extremely vindictive, then she may not even bother signing a contract with the other shop owners, because Sam will anticipate that she would be blinded by rage and cast aside her rational calculations in order to prosecute him out of vindictiveness.

If this suggestion is on the right track, then the anticipation of emotional consequences of actions (e.g. guilt aversion, Charness and Dufwenberg, 2006) might serve as a heuristic for assessing the reliability of commitments[10]. One consequence of this would be that individuals who do not experience moral emotions in a typical manner or do not understand them as others in their culture do, may also exhibit an anomalous understanding of commitments. In the context of development, this would imply that children's understanding of commitments should depend upon the development of their ability to anticipate moral emotions. Let us briefly consider some data that bears upon this conjecture.

First of all, the predominant view in developmental psychology is that children begin to exhibit pride and embarrassment around their second birthdays, showing public elation when performing well at difficult tasks, and blushing and hiding their faces when they do not do well at some task or other. It is noteworthy that this is around the time when they first pass the

[10] This is consistent with the idea, recently articulated by Szigeti (2013), that moral emotions serve as heuristics for assessing the moral status of actions.

mirror test (Bischof-Köhler et al. 1991), given that these emotions depend upon a self-other distinction and an understanding of how one appears from the outside, i.e. to the gaze of other people. As Rochat (2008) notes: 'Placed in front of a mirror with a mark on the face, the child often will not simply self-refer and remove the mark, but also show embarrassment, even blushing…In such secondary or self-conscious emotions, children demonstrate unambiguously that what they hold as representation of themselves (i.e. self-knowledge) factors the view of others' (249). But, of course, exhibiting or experiencing such emotions is different from understanding or anticipating them. And this is consistent with the finding that an understanding of complex moral emotions, such as guilt, pride and shame, continues to undergo fundamental development until at least around 7 or 8 (Harris 1989; Harris et al 1987; Nunner-Winkler and Sodian, 1988). Interestingly, children under this age rarely refer to such complex emotions in their speech (Ridgeway, Water and Kuczaj, 1985), and when presented with vignettes where an agent either succeeds or fails at some action with a moral significance according to their effort, their luck, or outside intervention, children younger than 7 or 8 are not proficient at inferring the resultant moral emotions, such as shame, guilt, pride and anger (Thompson, 1987; Thompson & Paris, 1981; Weiner, Graham, Ste and Lawson, 1982). Barden et al (1980) also reported that 4-5 year-olds predicted that a person would be 'happy' if they had committed an immoral act but not been caught, whereas 9-10 year-olds predicted that they would be scared or sad. When asked to predict their own emotions in such a situations, the children exhibited the same pattern. If the anticipation of moral emotions serves as an important heuristic for tracking commitments, then we should expect that children under about 9 should have difficulties in some cases – in particular in making judgments about violations of commitments.

Given this general timetable for the development of moral emotions and the ability to understand and anticipate them, we should expect children to begin to act in accordance with commitments, and to protest when others fail to, around their second birthdays. But we should not expect them to reliably anticipate whether people are likely to honor commitments, or to make reliable judgments about commitment violations, until around 9.

And indeed this pattern is strikingly confirmed by the existing data. First of all, Warneken and colleagues (2006) found that children as young as 18 months would protest when an experimenter with whom they were engaged in a simple joint action abruptly disengaged, thereby reneging on an implicit commitment to remain engaged until both parties to the joint action were satisfied that it had been completed. In a follow-up to this study, Gräfenhain and colleagues (2009) introduced a distinction between a condition in which the

experimenter made an explicit commitment to the joint action and a condition in which she simply entered into the action without making any commitment. Their finding was that 3 year-olds, but not 2 year-olds, protested significantly more when a commitment had been violated than when there had been no commitment. Moreover, in experiment 2 of the same study, the tables were turned and the children presented with an enticing outside option which tempted *them* to abandon the joint action. The finding here was that the children were less likely to succumb to the temptation if a commitment had been made, and in cases in which they did succumb, they were more likely to 'take leave', to look back at the experimenter nervously, or to return after a brief absence. These findings appear to suggest that they understand something about commitments by around 3[11].

Note, however, that the children in this study did not have to make judgments about commitments or anticipate future behavior. Indeed, the children frequently did first abandon the joint action and then return to it, demonstrating that they had failed to anticipate the negative emotional consequences of doing so. Presumably, they learn this gradually in the coming years and become much more proficient at making decisions that will not lead to negative emotional outcomes. This may take quite some time, however, as is demonstrated by two other studies probing children's understanding of commitments. In one of these studies, Mant and Perner (1988) presented children with vignettes in which one agent expects to meet a second agent at a particular time and place, and is disappointed that the second agent does not show up. In one condition (the commitment condition), the two agents had actually agreed to meet, while in the other conditions there was no such agreement (no-commitment condition). The children in the study, ranging from 5 to 10 years of age, were then asked to rate how naughty each character was. The finding was that only the oldest children (with a mean age of 9.5), and not the younger children, differentiated between the commitment condition and the no-commitment condition in rating the second agent's level of naughtiness. This may sound surprisingly late, but it is consistent with the findings of Astington (1988), who reported that children under 9 did not understand that one can only promise to bring about events over which one has some control, and that children as old as 11-13 judged that a speaker had not made a promise at all in cases in which the promise was unfulfilled.

In sum, the developmental findings provide support for the conjecture that moral emotions and the anticipation of moral emotions play important roles in the understanding of commitment. Once children experience moral emotions, they begin to act in accordance with

[11] A study by Hamann et al (2012) on 3-year-olds' understanding of commitment corroborates this point.

commitments and to protest when others fail to. But it is only later, once they are able to anticipate and make appropriate judgments about these emotions, that they are able to reliably anticipate and understand commitment-related behavior.

There is a problem, though: while moral emotions and sentiments may serve to make commitments more credible than they would otherwise be, it is also clear that they could stabilize cooperative behavior and deter cheaters even in the absence of any explicit commitment. Thus, it seems that both group identification and moral emotions and sentiments involve the engagement of automatic, intuitive processes that reduce motivational uncertainty with or without explicit commitments. The group identification route achieves uncertainty reduction by raising the salience of group interests compared to self-interests, while the emotional route does so by altering individuals' incentives and using emotions as a counter-weight to narrow self-interest.

8. Why bother with commitment?

The upshot of the discussion so far has been largely negative: although commitments may initially appear well-placed to serve the important function of reducing (motivational, instrumental and common ground) uncertainty within joint action, it is on the one hand not clear that they can be sufficiently credible to achieve this function, and on the other hand there are a host of other processes which do seem to contribute to the reduction of uncertainty. Why bother with commitment at all, then? In this section, we will argue that commitments may contribute to uncertainty reduction in important ways after all – and that they may do so in virtue of the types of process sketched in sections 6 and 7. Let us now consider several ways in which this may work.

First of all, commitments may trigger these processes when they would not already be triggered by circumstances. The most straightforward way in which this could occur is when there are no expectations, or no reliable expectations, to begin with. The public act of announcing a commitment can generate expectations by virtue of engaging moral sentiments in the manner discussed in section 6. For example, if an agent violates a commitment which she has made publicly, this may cause her to feel ashamed or guilty, and it may cause others to become angry or contemptuous. And since she and everyone else anticipate these emotional consequences, and everyone knows that they are undesirable outcomes which she is motivated to avoid, her commitment is credible, so she succeeds in generating expectations. To illustrate this, recall our example of Barbara the shop-owner and Sam the would-be

umbrella thief. Barbara may be able to make a credible commitment (to the other shop-owners or to the public) to prosecute all thieves even without signing a contract, and even if she is not known to be vindictive. This is because she and everyone else knows that if she violates her commitment, people will be angry or disappointed, and that she will want to avoid this outcome. It is worth pointing out that even in the absence of negative emotional outcomes of violating commitments, the risk of damage to one's reputation is also a strong motivating factor in favor of honoring commitments. But the likelihood of negative emotional outcomes if one fails to honor a commitment may well serve to enhance this motivation. Moreover, it may serve as a useful heuristic for assessing the likelihood of reputation loss, and the avoidance of negative moral emotions may even be an important proximal mechanism for reputation management.

Importantly, the notion of reputation also points towards a second way in which commitments can reduce uncertainty, namely by making existing expectations more salient and thereby also more reliable. How might this work? It could work if one proximal mechanism for managing one's reputation is to have a preference to fulfill others' expectations about when one will contribute to their goals. (Note that this would be far more efficient than having a preference for contributing to others' goals in general). Such a preference, if it exists, would be triggered whenever others' expectations become salient. In other words, commitments may engage a default preference to fulfill others' expectations by making those expectations salient.

There is some evidence that supports this conjecture. First of all, it is one possible explanation of the finding that people behave more generously in economic games when images of faces or eyes are present (Francey and Bergmüller 2012; cf. also Batson et al. 2006). It is also a plausible explanation of the robust finding that people tend to give away money in anonymous one-short dictator games (i.e. when an experimenter seems to expect them to) but do not just go around handing out money in everyday life (Camerer, 2003). This suggestion fits well with the findings from a classic study by Gaertner et al. (1973), in which a confederate called people on the telephone asking for money to help him out of a difficult situation. Political liberals were more likely to help than political conservatives – but only if they stayed on the phone long enough to hear his request, and in fact liberals were more likely to hang up sooner. These findings support two important claims: first of all, that people have a tendency to feel pressured into fulfilling others' expectations; and secondly, that they accordingly try to avoid learning of others' expectations in order to avoid being pressured into carrying out actions they do not want to carry out. More recently, Dana and colleagues

(2006) designed a dictator game in which the participant playing the role of dictator could pay $1 in order to exit from the dictator game, i.e. accepting a $9 payoff instead of being in a situation in which they could choose either to keep $10 for themselves or to give away as much as they wanted to. Many of the participants did indeed choose this option, but not in a condition in which they were told that the other person (the receiver) was unaware that she was a potential receiver in a dictator game. This suggests that making people aware of others' expectations makes them more likely to be cooperative. But does it reduce uncertainty? In other words, would one person be more confident that another person would cooperate with her if she could make her expectations known to him? To our knowledge, there is no data that bears directly on this question, but it could be tested by, for example, offering the receiver in a dictator game an exit option (e.g. $2) either privately or publicly (i.e. such that the dictator is aware of it). We would predict that receivers would be more likely to refuse such an exit option if the dictator were aware of it[12]. Indeed we would also predict that dictators would be willing to pay some amount in order to prevent the receiver's decision being common knowledge, i.e. to strategically avoid being confronted with others' salient expectations.

Thirdly, commitments may also reduce instrumental and common ground uncertainty by making people's behavior mutually salient and thereby engaging the kinds of automatic processes canvassed in section 6. For example, if Phil and Susan make a joint commitment to keep the house tidy, Phil may find himself observing Susan more than otherwise – perhaps to monitor and confirm that the joint commitment is being honored, or because he has been growing a bit lazy and hopes that she has too so that he does not need to feel guilty, or perhaps simply because he is pleased that the house is tidy and he takes pleasure in seeing it. In any case, by attending to each other's movements and thinking about each other, they are likely to trigger the full gamut of alignment processes referred to in section 6.

Fourthly, in addition to commitments reducing uncertainty by engaging and integrating a motley of automatic processes which do not depend upon or presuppose commitment, the converse relation may also sometimes obtain: automatic processes that can be engaged independently of commitments may also reduce uncertainty down to a point where people become willing to regard commitments as credible, to rely on them, and accordingly also to enter into interdependent commitments. If so, automatic processes could contribute to establishing the minimal common ground needed to induce us to form joint commitments.

[12] This idea was suggested in conversation by Christophe Heintz.

Fifthly, the various processes referred to in section 5 and 6 will tend to ensure that commitments are usually honored, and will thus build up an expectation that commitments just generally are honored. In other words, they will contribute to the establishment of trust, which leads people to rely on commitments even without thinking about whether the agent making the commitment is appropriately motivated in any given case to honor her commitment. The more one knows the person making the commitment, and the more experiences one has had with her in which she has not disappointed one's expectations, the stronger this default trust will tend to be. Indeed, it may be that the effects of perceived interpersonal closeness and perceived interpersonal similarity upon cooperation (Cornelissen et al. 2011) are driven at least in part by the general establishment of trust.

9. Conclusion

One way to conceptualize our endpoint is as a qualified rejection of the idea that cognitive dis-alignment and mutual distrust are the default assumptions at the background of human sociality. The various processes discussed in sections 6 and 7 constitute a background of default cognitive alignment and trust – in particular when people's situations, goals, expectations and movements are mutually salient. This background may help to explain why commitment, as Bratman and Gilbert agree, is such a prevalent feature of joint action in humans. We must emphasize that in speaking of default trust we do not want to go so far as to speak of general prosocial tendencies. Rather, the idea is that there are delimited ranges of cases in which the default expectation is cooperation. This explains, on the one hand, why commitments are so often credible. On the other hand, it also provides insight into how commitments fulfill their function of uncertainty reduction: they can do so by engaging and by integrating a motley of automatic processes which do not depend upon or presuppose commitment. Thus, joint actions would typically involve a mix of voluntary commitment processes, whether powered by practical rationality or by social normativity, and more basic, intuitive and automatic cognitive and emotional alignment processes. In addition, there can also be delimited ranges of cases in which the existence of a background of default cognitive alignment and trust is enough to lead people to engage in joint action, consistent with Alonso's suggestion (Alonso, 2009) that commitments can arise through joint action rather than being necessary prerequisites to joint action.

Acknowledgements

John Michael's work was supported by a Marie Curie Intra-European Fellowship (grant nr: 331140) within the framework FP7-PEOPLE-2012-IEF. Elisabeth Pacherie's work was supported by grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*.

References

Alonso, F. M. (2009). Shared Intention, Reliance, and Interpersonal Obligations*. *Ethics*, 119(3), 444-475.

Astington, J. W. (1988). Children's understanding of the speech act of promising. *Journal of Child Language*, 15(1), 157-173.

Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: the joint SNARC effect. *Social Neuroscience*, 3(3-4), 410-420.

AUTHOR (2012).

Axelrod, R (1984) *The Evolution of Cooperation*. New-York, NY: Basic Books

Bacharach, M. (2006). *Beyond Individual Choice*. N. Gold and R. Sugden (Eds.). Princeton: Princeton University Press.

Barden, R. C., Zelko, F. A., Duncan, S. W., & Masters, J. C. (1980). Children's consensual knowledge about the experiential determinants of emotion. *Journal of Personality and Social Psychology*, 39(5), 968.

Bateson M, Nettle D, Roberts G (2006) Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* 2: 412–414.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1), 27-52.

Bischof-Köhler, D. 1991. 'The Development of Empathy in Infants.' In *Infant Development: Perspectives from German Speaking Countries*, ed. M. E. Lamb and H. Keller, 245–73. Hillsdale, NJ: Erlbaum.

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology*: General,142(1), 143.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review*, 101(2), 327-41.

Bratman, M.E. (1999). *Faces of Intention: Selected Essays on Intentions and Agency*. Cambridge University Press.

Bratman, M. E. (2009b). Modest Sociality and the Distinctiveness of Intention. *Philosophical Studies*, 144, 149-165.

Brewer, M.B. (2003). Optimal Distinctiveness, Social Identity, and the Self. In M. Leary and J. Tangney (Eds.), *Handbook of Self and Identity*. (pp 480–491). New-York: Guilford Press.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Clark, H. H. (2006). Social actions, social commitments. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and interaction* (pp. 126–150). Oxford, New York: Berg.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. *Econometrica* 74, 1579–1601.

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are Social Value Orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin*, 37(8), 1080-1090.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2), 193-201.

de Bruijn, E. R., de Lange, F. P., Von Cramon, D. Y., & Ullsperger, M. (2009). When errors are rewarding. *The Journal of Neuroscience*, 29(39), 12183-12186.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (1998). 'Prosocial Development.' In: Eisenberg, N. (Ed.), *Handbook of Child Psychology*, vol. 3, Social, Emotional, and Personality Development, 5th ed., 701–78. New York: Wiley.

Francey D, Bergmüller R (2012) Images of Eyes Enhance Investments in a Real-Life Public Good. *PLoS ONE* 7(5): e37397. doi:10.1371/journal.pone.0037397

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New-York: WW Norton & Co.

Gaertner, S. L. (1973). Helping behavior and racial discrimination among Liberals and Conservatives. *Journal of Personality and Social Psychology*, 25, 335–341.

Gilbert, M. (1992). *On social facts*. Princeton: Princeton University Press.

Gilbert, M. (2006). Rationality in collective action. *Philosophy of the social sciences*, 36(1), 3-17.

Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies*, 144, 167–187.

Gold, N. and R. Sugden (2007) Collective Intentions and Team Agency. *Journal of Philosophy*, 104(3), 109-37.

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*,45(5), 1430.

Hamann, K., Warneken, F., & Tomasello, M. (2012). Children's developing commitments to joint goals. *Child development*, 83(1), 137-145.

Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Oxford: Blackwell.

Helm, Bennett W. 2008. Plural agents. *Nous* 42(1): 17–49.

Hepach, R., Vaish, A., & Tomasello, M. (2012). Young children are intrinsically motivated to see others helped. *Psychological Science*, 23(9), 967-972.

Hume, D., 1740 [1888 1976], *A Treatise of Human Nature*, L. A. Selby-Bigge (ed.), rev. 2nd. Ed. P. H. Nidditch (ed.), Oxford: Clarendon Press.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*,330(6012), 1830-1834.

Kutz, C. (2000). Acting Together. *Philosophy and Phenomenal Research*, 61, 1, 1-31.

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

Mant, C. M., & Perner, J. (1988). The child's understanding of commitment. *Developmental Psychology*, 24(3), 343.

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child development*, 1323-1338.

Pacherie, E. (2011). Framing Joint Action. *Review of Philosophy and Psychology*, 2(2): 173-192.

Pacherie, E. (2013). Intentional joint agency: shared intention lite. *Synthese,* 190, 10: 1817-1839

Pezzulo, G. (2011). Shared representations as coordination tools for interaction. *Review of Philosophy and Psychology*, 2(2), 303-333.

Pezzulo, G., & Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research*,211(3-4), 613-630.

Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, 17(8), 413.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416), 427-430.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of unintentional and intentional interpersonal coordination. *Human Movement Science*, 26, 867–891.

Ridgeway, D., Waters, E., & Kuczaj, S. A. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology*, 21(5), 901.

Roch, S. G., Lane, J. A., Samuelson, C. D., Allison, S. T., & Dent, J. L. (2000). Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups. *Organizational Behavior and Human Decision Processes*, 83(2), 185-212.

Rochat, P. (2008). 'Know Thyself!'...But what, how and why?. In Sani, F. (Ed.), *Individual and Collective Self-Continuity*: Psychological Perspectives. Lawrence Erlbaum Publishers

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2012). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*.

Searle, J. (1990) Collective Intentions and Actions. In P.Cohen, J. Morgan, and M.E. Pollack (Eds.), *Intentions in Communication* (pp. 401-416). Cambridge, MA: Bradford Books, MIT Press.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting Stimulus–Response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1234–1246.

Shpall, S. (2014). Moral and rational commitment. *Philosophy and Phenomenological Research*, 88(1), 146-172.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition*. Oxford: Blackwell.

Szigeti, A. (2013). No Need to Get Emotional? Emotions and Heuristics. *Ethical theory and moral practice*, 16(4), 845-862.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96-102

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge: Cambridge University Press.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), 149–178.

Thompson, R. A. (1987). Empathy and emotional understanding: The early development of empathy. In: Eisenberg, N and Strayer, J. (Eds.) *Empathy and its development* (pp.119-145). Cambridge University Press.

Thompson, R.A. (1987).Development of children's inferences of the emotions of others. *Developmental Psychology,* 23, 124-131.

Thompson, R.A., and Paris, S.G. (1981). Children's inferences about the emotions of others. Paper presented at the biannual meeting of the Society for Research in Child Development, Boston.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, 52(3), 314-320.

Tsai, C. C., Kuo, W. J., Hung, D. L., & Tzeng, O. J. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, 20(11), 2015-2024.

Tuomela, R. (2007). *The philosophy of sociality*. Oxford: Oxford University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology*, 46(4), 693-695.

van der Wel, R. P., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, 130(1), 128-133.

van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage*, 43(4), 808-814.

van Schie, H. T., Mars, R. B., Coles, M. G., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, 7(5), 549-554.

Vesper, C., S. Butterfill, N. Sebanz, and G. Knoblich. (2010). A minimal architecture for joint action. *Neural Networks,* 23(8/9), 998–1003.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, 77(3), 640-663.

Weiner, B. Graham, S, Stern P., Lawson, M.E. (1982). Using affecting cues to infer causal thoughts. *Developmental Psychology*,18, 278-286.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3), 460.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental psychology*, 28(1), 126.