The Developmental Origins of Commitment

John Michael*

Philosophy Department, University of Warwick

& Department of Cognitive Science, Central European University

&

Marcell Székely

Department of Cognitive Science, Central European University

Forthcoming in the *Journal of Social Philosophy*

Accepted 8 January 2018

Abstract: As adults, we are quite proficient in generating commitments, and in identifying, keeping track of, and responding appropriately to our own and others' commitments. This proficiency is fundamentally important for uniquely human forms of sociality. By illuminating the cognitive processes underpinning commitments, we may therefore gain insight into the ways in which human cognition is unique, and into the ways in which it is shared with other species. In pursuing this aim, one valuable strategy is to investigate the emergence of an understanding of commitment in ontogeny, i.e. to isolate distinct components of this proficiency as they emerge, and to learn how they relate to each other, which are the most basic, etc. Our aim in this paper is to contribute to this project by articulating a theoretical framework to structure research on the emergence of an understanding of commitment in childhood.

Keywords: Commitment, Development, Joint Action, Cooperation, Normative Protest

* Corresponding Author: j.michael.2@warwick.ac.uk

1. Introduction

Commitments are a core feature of human social life. They make individuals' behaviour predictable in the face of fluctuations in their desires and interests, thereby facilitating the planning and coordination of joint actions involving multiple agents. Moreover, they make people willing to perform actions that they would not otherwise perform. For example, an employee performs her job every day because her employer has made a credible commitment to pay her at the end of the month.

As adults, then, we appear to be quite proficient in generating commitments[1], and in identifying, keeping track of, and responding appropriately to our own and others' commitments. Moreover, this proficiency appears to be fundamentally important for uniquely human forms of sociality. By illuminating the cognitive processes underpinning this proficiency, we may therefore gain insight into the ways in which human cognition is unique, and into the ways in which it is shared with other species. In pursuing this aim, one valuable strategy is to investigate the emergence of an understanding of commitment in ontogeny, i.e. to isolate distinct components of this proficiency as they emerge, and to learn how they relate to each other, which are the most basic, etc. Our aim in this paper is to contribute to this project by articulating a theoretical framework to structure further research on the emergence of an understanding of commitment in childhood. Our question, then, is: *How do children attain a mature proficiency at identifying, keeping track of, and responding appropriately to their own and others' commitments?*

At first blush, it may seem that there is a simple answer to this question: children acquire the concept of commitment sometime during development, and it is the mastery of this concept which underpins adults' proficiency in generating commitments, and in identifying, keeping track of, and responding appropriately to one's own and others' commitments. In the following section (section 2), we will evaluate this simple answer and identify theoretical and empirical reasons for finding it unsatisfactory. In section 3, the main body of the paper, we articulate and defend the hypothesis that the aforementioned proficiency rests upon an intuitive sense of commitment, which is more basic than a conceptual understanding of commitment, and which the latter builds upon and extends. In

[1] For simplicity, we will use the term 'commitment' to refer to interpersonal commitment, i.e. for the purposes of this paper, we are not considering commitments that an individual makes to herself.

section 4, we offer some speculations about the relationship between the sense of commitment and the concept of commitment. In section 5 we conclude by returning to our question about the origins of characteristically human proficiency in managing commitments.

2. A Simple Conjecture

2.1 Do children gain proficiency by acquiring the concept of commitment?

So far, we have been using the term 'commitment' loosely, without specifying precisely what we mean by it. At this point, it will be useful to offer a definition. This will help us to establish as clearly as possible what the explanandum is that we are aiming to account for.

According to a conception that is standard within the philosophical literature, a commitment is a relation among a committed agent, a second agent to whom the commitment has been made, and an action which the committed agent is obligated to perform. The committed agent is obligated to performing the action because she has given an assurance to the second agent that she will do so, and the second agent has acknowledged this under conditions of common knowledge (Gilbert, 2009; Searle, 1969; Scanlon, 1998; Shpall, 2014).

This standard conception presents a clear characterisation of paradigm cases of commitments arising through promises or other explicit verbal assurances. We will call the concept picked out by the standard conception 'commitment in the strict sense'. In so doing, we do not prejudge the question as to whether commitment in the strict sense can also be applied to cases of implicit commitment (although we are skeptical about this). In other words, the term 'commitment in the strict sense', while clearly tailored to cases of explicit commitment, is not synonymous with 'explicit commitment'.

The standard conception provides a straightforward explanation of why adult humans expect others to perform actions they are committed to performing, and are prepared to rely on those expectations: commitments give rise to obligations, and people are entitled to expect (in a normative sense) others to do what they are obligated to do. Of course people don't always do what they are obligated to do, and we don't always expect them to (in a non-normative sense). But we do generally[2] take ourselves to be entitled to censure them if they don't.

[2] There are exceptions -- for instance, there may be a competing obligation that takes precedence.

A simple conjecture about how children acquire proficiency with commitments is that they acquire the concept of commitment in the strict sense. According to this conjecture, possession of the concept of commitment in the strict sense leads children to act in accordance with their commitments and to otherwise acknowledge the appropriateness of censure, and to believe that they themselves are entitled to censure others who do not act in accordance with their commitments. But, though acquiring the concept of commitment in the strict sense is surely very important, there are compelling reasons to be unsatisfied with this simple conjecture. We will first identify theoretical reasons (subsection 2.2), and then turn to empirical considerations (subsection 2.3) which also compel us to look beyond the simple conjecture.

2.2 Theoretical reasons for being unsatisfied with the simple conjecture

There are numerous features of our mature human proficiency in managing commitments that are not yet explained by appealing to the concept of commitment in the strict sense. Specifically, the concept does not clarify (a) how people determine when commitments are in place in the absence of an explicit agreement or promise, (b) how they determine what the precise content of an explicit or implicit commitment is, (c) how they assess the appropriate degree of commitment, and (d) how they determine what grounds are acceptable for abandoning a commitment.

Consider the following example: Roger often volunteers as an assistant at a local retirement community. One of the residents, Patricia, is celebrating her birthday today. Roger was not explicitly invited, but he knows that Patricia would be delighted if he dropped by, and that the other people involved could use his help setting up for the party, ensuring that it runs smoothly, and cleaning up afterward. He may not have made any explicit commitment to anyone, but he may nevertheless have a sense that he is implicitly committed, either to Patricia, or to the other people involved, and this may motivate him to attend the party and to help out anyway (See (a) above). Or he may have agreed to drop by but be surprised to discover that he is expected to help out by parking cars for the guests (See (b) above). Or he may even have agreed to help park the cars but be surprised to discover that he is in fact expected to persist at this cheerless task for several hours in the hot sun (See (c) above). Or, if we tweak the example slightly, we might also imagine that he did agree to go and help park the cars, but that he would now like to get out of this commitment because his friend has

invited him to go to the pub for drinks. On the face of it, this excuse does not seem to be very compelling. But what if he suspects that his friend has been depressed and that it is important to him that they discuss something together?

Such cases are very common in everyday life. But the concept of commitment in the strict sense will not on its own be sufficient to make an appropriate judgment in such cases. This is because the concept of commitment in the strict sense provides no reason for Roger to show up to the party at all if he has not expressed his willingness to do so to any relevant party under conditions of common knowledge, and even less reason to park cars for the guests — and yet, a mature adult would often feel committed and act accordingly in such cases, and expect others to do so as well. Nor does the concept of commitment in the strict sense help in deciding which grounds for abandoning a commitment are appropriate and which are not. This means that in developing a mature proficiency in managing commitments, it is not sufficient for children to acquire the concept of commitment in the strict sense.

2.3 Empirical reasons for being unsatisfied with the simple conjecture

In order to evaluate the empirical credentials of the simple conjecture, it will be necessary to begin by specifying the predictions that it generates. Of course, the simple conjecture is very broad as we have formulated it. As a result, it does not entail very many specific predictions about issues for which we would in fact like to have specific predictions. For example, it does not entail any specific positive predictions about when children will acquire the concept of commitment in the strict sense, although it does of course predict that they will not acquire this concept before acquiring the other concepts of which it is composed, such as the concepts of 'obligation' and 'common knowledge', and possibly also the concepts of 'intention', 'belief' and 'desire,' which feature indirectly in the definition. There is evidence that one year-olds are able to identify intentions (Behne et al., 2005). At the moment, however, it is unclear when children are able to understand the concepts of (Butterfill & Apperly, 2013; Christensen and Michael, 2016; Carruthers, 2013), desire (Rakoczy, 2007; Steglich-Petersen & Michael, 2015), obligation (Astington, 1988; Rakoczy et al., 2008; Vaish et al., 2011), and common knowledge (Carpenter & Liebal, 2011). In view of this uncertainty, we will not evaluate this prediction here.

The simple view also entails that once children have acquired the concept, they will exhibit a suite of behavioral tendencies which are licensed by the concept. They should, for example, be inclined to wait for a partner to whom they are committed and who is slower than they are in the context of a activity, to check on the progress of their partner(s), to offer help where appropriate, to refrain from abandoning the activity until all parties are satisfied that the goal has been achieved or until all have agreed to abandon it (Gilbert, 1989; Tuomela, 2007). They should also be inclined to censure others who violate explicit verbal agreements to perform actions, and acknowledge other' rights to censure *them* if they themselves do so (Gilbert, 1989). This may take the form of explicitly censuring and explicitly acknowledging others' right to censure, or it might take a more implicit form. For example, they may be inclined to cry and/or to protest if agreements are violated, but without explicitly stating the reason why. Similarly, they may exhibit signs of guilt or of fearing punishment when they themselves violate agreements. The crucial point is that the simple conjecture predicts that once children acquire the concept of commitment in the strict sense, there should be an uptick in these behaviors, because these behaviors are licensed by the commitments, as one would understand by grasping the concept. What do the data show?

Gräfenhain and colleagues (2009) implemented a paradigm in which an experimenter and a child play various games together. In Experiment 1 of their study, they were interested in how children would react when, at some point, the experimenter abruptly stopped playing. Specifically, they compared a condition in which the experimenter had made an explicit commitment to the joint action and a condition in which she had simply entered into the action without making any commitment. Interestingly, 3-year-olds, but not 2-year-olds, protested significantly more when a commitment had been violated than when there had been no commitment. In Experiment 2 of the same study, the tables were turned and the children were presented with an enticing outside option that tempted *them* to abandon the joint action. The children were less likely to succumb to the temptation if a commitment had been made. 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.

The interpretation of these findings suggested by the simple conjecture is that children acquire the concept of commitment in the strict sense by around three. But consider a study conducted by Mant & Perner (1988), in which children were presented with vignettes describing two children on their way home from school, Peter and Fiona, who discuss whether to meet up and go swimming later on. In one condition, they make a joint commitment to meet at a certain time and place, but Peter decides not to go after all, and

Fiona winds up alone and disappointed. In the other condition, they do not make a joint commitment, because Fiona believes that her parents will not let her. She is then surprised that her parents do give her permission, and she goes to the swimming pool to meet Peter. In this condition, too, however, Peter decides not to go after all, so again Fiona winds up alone and disappointed. 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) judged Peter to be more naughty in the commitment condition than in the no-commitment condition. This may seem late, but it is in fact consistent with the findings of a study by Astington (1988), who reported that children under 9 fail to understand the conditions under which the speech act of promising gives rise to commitments. If we take these results at face value, it suggests that children do not master the concept of commitment until they are much older than the children in Gräfenhain and colleagues' (2009) study. This indicates that we need some other explanation of the pattern observed with these younger children.

More generally, the simple conjecture does not provide us with any guidance in generating predictions about what components of the concept of commitment may emerge first, or about what behavioral tendencies may emerge first (waiting for a partner, checking on her, helping her, persisting until all parties are satisfied that the goal has been reached, protesting if a partner abandons a joint action, etc.). In other words, the simple conjecture presents a complex concept and a suite of behaviours licensed by the concept as a single package. But these components may come apart, and some may be more basic than others. The simple conjecture does not tell us in what order these components should emerge, which components are most basic, or how the developmental process should unfold.

Moreover, there is a further detail in the findings reported by Gräfenhain and colleagues which should give us pause. Specifically, it is not the case that the two-year-olds do not protest at all, and only the three-year-olds understand the situation well enough to feel entitled to protest. In fact, there is no increase in appropriate normative protest from two to three. On the contrary, the two-year-olds protest just as much in both conditions as the three-year-olds do in the commitment condition. This suggests that the sense of entitlement that inspires protest over an unfulfilled expectation is not the product of developmental changes over the third year but, rather, it is the default that is already in place by two or earlier. What changes in the third year is that children learn that they are not always entitled to expect contributions to their goals. In other words, the developmental process chips away from, rather than adding to, the cognitive architecture that underlies the protest behavior.

There is also a further detail in Mant & Perner's (1988) study which bears emphasizing: 22 of the 46 6-year-olds actually rated the protagonist as being naughty in both conditions (while 11 rated him as neutral in both conditions), i.e., when Peter had violated a commitment and thereby caused Fiona to be disappointed and sad, and when he had not made any commitment in the first place and Fiona had been disappointed and sad. It is as though, whenever a goal is not achieved and somebody is left disappointed, the default is to assign blame, and to work out the details later. This is not the pattern that one would expect on the basis of the simple conjecture. This is because the simple conjecture predicts that normative protest emerges as a result of the understanding that one is entitled to protest because a commitment in the strict sense is in place.

We propose to develop a different approach to explaining the developmental trajectory of children's proficiency in identifying, keeping track of, and responding appropriately to our own and others' commitments. Rather than taking the concept of commitment in the strict sense as a starting point, and interpreting the findings of Gräfenhain and colleagues (2009; 2013; cf. also Hamann et al., 2012) as evidence that three-year-olds understand and respond to commitments in the strict sense, we will attempt to identify a broader, less complex phenomenon that young children may understand and respond to even in the absence of a sophisticated understanding of common knowledge, obligations and the speech act of promising. Our aim will be to explain how an understanding of commitments emerges through engagement in joint actions, as several distinct cognitive and affective mechanisms are integrated and calibrated through social experience. Our more psychological approach (i.e. in contrast to an approach based on normative notions) resonates with the view of many theorists that a simplified conception of joint action is needed in order to account for young children's engagement in joint actions (Butterfill, 2012; Brownell, 2006; Tollefsen, 2005).

3. A Minimal Approach

3.1 Conceptualizing the sense of commitment

In theorizing about the 'broader, less complex phenomenon' that children are progressively able to identify and respond to, we will draw upon Michael, Sebanz & Knoblich's (2015)

characterization of the minimal structure of situations in which a subjective *sense of commitment* can arise. They characterize the minimal structure as follows:

(i) There is an outcome which an agent (ME) either desires to come about, or which is the goal of an action which ME is currently performing or intends to perform. We will refer to this outcome as 'G' (for 'goal').

(ii) The external contribution (X) of a second agent (YOU) is crucial to bringing about G.

Clearly, conditions (i) and (ii) specify a broader category than that of commitment in the strict sense. Nevertheless, situations with this structure may elicit a sense of commitment on the part of one or both agents. We stipulate the following working definition of the sense of commitment:

ME has a sense that YOU is committed to performing X to the extent that ME expects X to occur because (i) and (ii) obtain.

YOU has a sense of being committed to performing X to the extent that YOU is motivated by her belief that ME expects her to contribute X.

Clearly, conditions (i) and (ii) specify a broader category than that of commitment in the strict sense. In particular, while commitments in the strict sense arise intentionally (Gilbert, 1989), an agent can come to have a sense of commitment to doing X as a side effect of an intentional action. For example, Sam is cleaning up the living room and picks up a ball that had been lying on the floor. As it happens, his dog Woofer notices this and bounds over to him, apparently ready to play fetch. Sam was not intending to play fetch and does not particularly desire to, but may now feel obliged to, because he has generated an expectation on the part of Woofer that they will now play fetch together. Thus the unintentional generation of expectations can lead individuals to sense that a commitment is in place. Of course, if Sam intentionally makes eye contact with Woofer and waves the ball around in the air, he thereby generates a high degree of commitment to playing fetch. And if Woofer is sensitive to these cues, they may lead him to have a high expectation that Sam is now going to play fetch with him.

So far, then, we have characterized the sense of commitment in terms of agents expecting external contributions (i.e., X) to be made because the minimal structure is in place [i.e., conditions (i) and (ii)], and/or being motivated to make contributions because they believe they are expected to. Our proposal is that children first acquire a sense of commitment (as we have characterized it), and that this sense of commitment is gradually

calibrated through social experience to give rise to a mature proficiency in managing commitments. In order to spell out this proposal, we will first need to explain how a sense of commitment would arise in the first place. In other words, why would children, or indeed anyone at all, have such expectations and/or motivations? Next, we will need to explain how the sense of commitment could develop into a mature proficiency in managing commitments.

Our attempt to meet these challenges will consist of three steps, which we will discuss in the next three subsections:

1) There are numerous mechanisms leading humans (and quite possibly in some cases other species as well) to be motivated to contribute X in situations in which the minimal structure is instantiated (i.e. (i) and (ii) obtain), and some of these mechanisms are present already in infancy (Section 3.2).
2) There are numerous mechanisms leading humans (and in some cases other species as well) to expect X to occur because (i) and (ii) obtain (Section 3.3)
3) These expectations and motivations reinforce each other over time, and are calibrated through joint actions and other social experiences, leading children ultimately to a mature proficiency in identifying, keeping track of, and responding appropriately to our own and others' commitments (3.4)

3.2 How would YOU come to be motivated to do X because the minimal structure is instantiated?

It will be useful to differentiate two subtypes of the minimal structure outlined above, one based upon a rich conception of goals, and one based upon a lean conception of goals. Both of these subtypes sometimes obtain, and for each of them there are mechanisms triggering YOU's motivation to do X. What, then, is the difference between the rich and the lean conception of a goal?

At a bare minimum, a goal is an outcome of an agent's movements. But clearly this is not enough to distinguish goals from incidental consequences of movements. For example, stepping on and killing a bug may be a consequence of walking across the room, whereas the goal may be to place some books in the cabinet. Intuitively, an outcome of an action is only a goal of that action if the action is performed because it is likely to bring about that outcome.

There are various ways of articulating this idea. In particular, they differ with respect to whether or not they appeal to the mental representations of the agent carrying out the action. Butterfill & Apperly (2013), for example, offer a lean characterization of goals which avoids making appeal to the mental representations of the agent. They write:

> We stipulate that for an outcome, *g*, to be the goal of some bodily movements is for these bodily movements to occur in order to bring about *g*; that is, *g* is the function of this collection. Here "function" should be understood teleologically. On the simplest teleological construal of function, for an action to have the function of bringing about *g* would be for actions of this type to have brought about *g* in the past and for this action to occur in part because of this fact . . . The virtue of this way of representing goals is that it allows them to be inferred from actions without appealing to intentions, beliefs, preferences or other psychological states. (Butterfill & Apperly, 2013, p. 613)

This characterization (by design) eschews mentalistic talk of what an agent *intends* or *desires* to bring about, or is *trying* to bring about, or of what outcome the agent *represents*. Instead, it distinguishes the goal of an action from other outcomes of the action by appealing to the notion of a function (understood teleologically, cf. Millikan, 1984): the action is performed because on previous occasions performing the action led to the outcome. This characterization has the virtue of simplicity, and the absence of mentalistic language may well make it easier to operationalize.

On the other hand, it may be problematic in cases in which an action is performed for the first time, or where it is likely to lead to a different outcome than it has in the past. Moreover, the very same movements can function to bring about different outcomes in different situations, depending on features of the context, including various mental states of the observed agent (Jacob & Jeannerod 2005; Fiebich & Coltheart, 2015; Michael & Christensen, 2016). In order to address such cases, it may be useful to appeal to *intentions*, *expectations*, *desires*, *trying* or other mental *representations* that guide the action (Huang & Bargh, 2014; Aarts & Dijksterhuis, 2000). On such a richer view, an outcome of an action counts as a goal if the agent's actions are guided by a representation of that outcome. A representation of a particular outcome may, for example, make it possible to modify the action in light of feedback or of changing circumstances such as to increase the likelihood of efficiently bringing about the outcome.

The minimal structure may be instantiated in either form, with YOU identifying ME's goal in the lean sense or in the rich sense. Either way, this can lead YOU to be motivated to bring about G, as we will now explain.

Lean goals in the minimal structure

Can an agent identify the goals of actions without ascribing mental states to the agents of those actions? Csibra and Gergely (1998) proposed a computational description of such a mechanism, dubbed the 'teleological stance.' In their words: 'an action can be explained by a goal state if, and only if, it is seen as the most justifiable action towards that goal state that is available within the constraints of reality' (Csibra and Gergely 1998, p. 255). According to their account, an agent observing another agent's body movements identifies the other agent's goals from the pool of possible outcomes of the observed agents movements by excluding those outcomes for which there would be more efficient ways to achieve them. The remaining outcomes are the goals of the agent's actions.

Could then the second agent (YOU) in a case in which the minimal structure is instantiated apply the teleological stance to identify the first agent's (ME's) goal (G)? It seems that YOU could not. One limit of the teleological stance is that it requires successful actions as inputs for correctly computing the goals of actions because a failed action is usually not the most efficient way to achieve the outcome which is actually its goal. In our case, ME is directing his/her action at a goal (G) for which an external contribution of a second agent (YOU) is crucial to bringing about G. By definition, ME's actions towards G would fail unless YOU contributes. Consequently, YOU could not identify the goals of ME's actions in the minimal structure.

However, Butterfill (in preparation) argues that this limitation of the teleological stance can be overcome if the two agents are similar with respect to their ability to identify the most efficient actions to bring about outcomes. Agents don't need to be good at identifying the most efficient ways to bring about outcomes; what matters is that they rely on similar processes to compute the best available ways of achieving outcomes. When this requirement is fulfilled in the minimal structure, YOU could identify ME's goals of actions in the lean sense.

What representations and algorithms are involved in applying the Teleological Stance? Currently there are two hypotheses. Csibra and Gergely (1998, 2003, 2013) hypothesize that agents use the computational strategy of the Teleological Stance explicitly in their reasoning about the goals of actions. Sinigaglia and Butterfill (2016) hypothesize that

when an agent observes another agent's actions s/he often represents these actions motorically, and these motor representations trigger processes associated with performing actions, which in turn lead to expectations concerning the goals of actions (Motor Theory of Goal Tracking). To our knowledge, neither of the hypotheses generate incorrect predictions. However, Sinigaglia and Butterfill's hypothesis is better supported because it correctly predicts that impairing agents' abilities to represent actions motorically (by tying hands or by transcranial stimulation) can also impair goal tracking (Ambrosini, Sinigaglia and Costantini 2012; Costantini et al. 2013), and enhancing action abilities can lead to better goal tracking performance (Sommerville, Woodward and Needham 2005; Sommerville, Hildebrand and Crane 2008).

Whatever ways an agent (YOU) identifies the outcomes at which the other agent's (ME) actions are directed, there is still a need for a mechanism that explains why YOU would treat the identified goals as her own. Michael and Szekely (2017) propose such a mechanism, which they dub 'goal slippage'. On their account, goals that are identified in instances instantiating the minimal structure are sometimes represented as motor representations within the observer's motor system -- namely, when the observed action is in their own motor repertoire. When this occurs, the identified goal becomes the observer's own goals, and observer will automatically act to bring about the identified goals unless some other mechanisms inhibit their automatic action. For example, YOU may observe as ME attempts to toss a pillow onto a seat in the row in front of her on an airplane, and notice that the pillow, unbeknownst to ME, has rolled onto the floor. In such as case, YOU may pick up the pillow and place it on the seat in order to facilitate the achievement of the goal. Although an agent's motivation to bring about such goals may generally be lower than her motivation to bring about endogeneously generated goals, goal slippage could nevertheless increase the likelihood of YOU doing X.

Given that goal slippage is hypothesized to be an automatic process, Michael and Szekely (2017) suggest that it should be more likely to occur when executive resources are occupied (e.g. under cognitive load). This generates the prediction that spontaneous helping behavior (Warneken et al., 2006; Warneken & Tomasello, 2007; for discussion and further references, see Michael & Székely, 2017) should increase under cognitive load.

It is worth noting that YOU's motivation to do X in such cases is not an instance of the sense of commitment according to the definition we have adopted. This is because our definition applies only to cases in which the motivation to do X is triggered at least in part by the belief that a second agent (ME) expects one to do X. Nevertheless, it contributes to the

establishment of a default expectation which, as we shall see further below, is crucial for the sense of commitment.

Rich goals in the minimal structure

In cases in which the minimal structure is instantiated and YOU identifies ME's goal in the rich sense, there are many reasons why YOU might thereby be motivated to contribute X.

For example, YOU might enjoy pursuing goals together with others (*collectivity preference*). If so, YOU may take satisfaction in successfully coordinating with other passengers in order to get everyone to their seats. Or YOU might also be motivated by an *altruistic* preference for seeing others meet their goals (or for seeing ME specifically meet ME's goals, if YOU happens to know and like ME). For example, YOU might be especially motivated to assist an injured or elderly fellow passenger in taking their seat. A further possibility, following a hypothesis put forward by Heintz and colleagues (2015), is that YOU may be motivated to do X because she thinks that ME expects her to (*expectation fulfillment*). Indeed, insofar as YOU believes that ME expects X to occur, YOU may expect ME to show signs of conflict if X does not occur, and indeed to address YOU directly with these signs of conflict. For example, if the fellow passenger has tossed her book onto the window seat and then backed up into the aisle and cleared space for YOU to stand up and get out of her way, then YOU may infer that ME has a specific expectation about what YOU will do, and sense that the path of least resistance is to fulfill that expectation. A final possibility is that YOU may simply have an aversion to the signs of conflict that ME exhibits if the goal is not reached – for example if YOU is anxious to return to her newspaper and is annoyed by the disturbance created by the other passenger (*aversion to others' distress*).

Taken together, these factors may conspire to sustain a default preference on the part of YOU to contribute X when she detects that a situation with the minimal structure is in place.

3.3 How would ME come to expect YOU to do X because the minimal structure is instantiated?

We believe that there are numerous reasons why infants in the role of ME tend to expect YOU to do X in cases instantiating the minimal structure. At the most basic level, the expectation that G will occur when desired may have the status of a default in infants. This is

because an infant may not entertain the possibility that G is only her own goal, or an outcome that only she desires to be brought about (Piaget 1950). A default expectation that G will occur when desired would be consistent with many experiences that infants and young children have in their first years of life. Indeed, as soon as infants begin pursuing goals, there is usually at least one parent who is motivated to support them in their goals. Moreover, infants experience distress or conflict when their goals are not met.

Our hypothesis is that this default expectation of G is progressively qualified over the course of development -- i.e. it becomes increasingly context-specific as children develop more sophisticated abilities to understand the instrumental structure of action, to evaluate agents, and to identify and integrate more and more relevant factors which are relevant to predicting whether X will occur. A first step beyond the very basic default expectation which we have proposed is to identify specific agents that are associated with successful outcomes. For example, an infant may come to associate mommy with good outcomes, and thus expect G to occur specifically when Mommy is present. A further important step is to be able to identify the specific external contributions (X) which are required for their goals (G). For example, Billy may come to notice that in order to bring about the goal of feeding him (G), Mommy needs to grasp the bottle and present it to his mouth (X). When Billy has attained this level of sophistication, his default assumption will be that those contributions (X) will be made. And instances in which he does not meet a goal because X is not contributed may also elicit signs of distress and/or conflict. Moreover, as Billy gets older, he will also need to learn to evaluate many more factors, such as whether Mommy is aware of his desire to eat, whether it is reasonable to expect her to feed him now (which would, for example, not be the case if Mommy is currently driving in heavy traffic), whether she has made a promise to feed him at this moment in particular, etc.

One possibility raised by this view of development is that this bedrock sense of entitlement remains into adulthood, usually below the surface of behavior. Indeed, we suspect that this is the case, and that this default attitude can be glimpsed in those moments when one is stressed or tired and, struggling to tie one's shoe or to close a drawer, catches oneself cursing at the shoe or the drawer and feeling inclined to mete out punishment to whatever objects or agents happen to be around. Our conjecture is that, psychologically, this sense of outrage and frustration is the very same sense of outrage and frustration as what one experiences when there really is an agent who is to blame for some normative violation.

Be that as it may, such a default expectation -- suitably qualified on the basis of knowledge gained through social experience -- could generate or reinforce specific

expectations that ME would not otherwise have about contributions (X) to be made to ME's goals or to outcomes which ME desires to be brought about (G).

But on top of this basic default expectation, there are many further reasons for ME to expect YOU to do X -- namely, the very same reasons why YOU is in fact often motivated to do X (as we set out in the previous section), namely because YOU is motivated by such mechanisms as altruism, a collectivity preference, an aversion to others' distress, and an aversion to disappointing other' expectations.

3.4 Expectations and motivations reinforce each other over development

In the previous two subsections, we gave reasons why some agents, in particular infants and young children, may expect X to occur because (i) and (ii) obtain. We also gave reasons why some agents, in particular infants and young children, may sometimes be motivated to contribute X because because (i) and (ii) obtain, and also sometimes because they believe that they are expected to. In this section, we will explain how these expectations and motivations can reinforce each other over the course of development, and how the sense of commitment can thereby become calibrated to the norms within a culture.

On the one hand, ME's default expectation that others (such as YOU) will contribute to ME's goals will be likely to be met and reinforced if other agents (such as YOU) are indeed likely to contribute because of the processes referred to in the previous two subsections. On the other hand, YOU will be more likely to contribute X if YOU believes that ME expects this.

This does not imply that children (or, for that matter, adult humans) always expect others to contribute X in situations instantiating the minimal structure, nor that they always contribute X when they think they are expected to. In many such instances in which an agent expects X, X simply does not occur. Indeed, even infants' and young children's parents don't always support their goals or fulfil their desires. So, as noted already above, in order to differentiate among various degrees of likelihood that X will occur, children must develop a more nuanced sensitivity to features of interactions that carry information about the reliability of various kinds of cues to X in various situations.

Is YOU aware of ME's expectation of X? Did YOU do anything to cause ME to have this expectation? If so, was this intentional? Is there any precedent for this expectation? That is, has YOU made the contribution of X in previous similar situations? If, for example,

Daddy has played catch with Leonardo every Saturday in the garden for many months, it is more reasonable to expect this to occur this Saturday than if Daddy had only done it once or twice. Similarly, it is also important to assess to what extent ME is relying on X for the achievement of G). If X is something that can really only be achieved with YOU's contribution, and if it is very important, then it is less appropriate for YOU to refuse unless there is a good reason. Leonardo, for example, can play with some of his toys alone if Mommy is busy, but his new wiffle ball bat is only fun to play with if someone pitches the wiffle ball for him to swing the bat at -- so it is more reasonable to expect Mommy to play together with him, and all the more so if he needs to practice for a wiffle ball game at his friend's birthday party the following day.

Moreover, through social experience over many years, children also learn when it is appropriate to abandon or postpone commitments. For example, if Daddy promises to take Leonardo to the zoo, but then has to rush off to work to deal with an urgent matter, Leonardo will need to understand that Daddy's urgent matter provides a good reason to postpone the zoo trip until the following day.

By the same token, it would be inefficient for an agent *always* to contribute to others' goals or desired outcomes whenever she believed that she were expected to. Hence, children must also learn to apply the same criteria in determining whether to make crucial contributions to others' goals or desired outcomes as they apply in determining whether to expect others to make those contributions. And more generally speaking, the processes which we have postulated as underpinning a sense of commitment are likely to become calibrated through experience to match those of other people in their culture, and to conform to cultural norms concerning when it is considered appropriate to make contributions to others' goals and to expect contributions from others. As a result, people's expectations about the extent to which others will be motivated by such processes will roughly match the extent to which others really are so motivated.

4: What about the Simple Conjecture?

So far, we have given an account of how various sources of motivation and of expectations reinforce each other over the course of development. Through this long process of mutual reinforcement, expectations are calibrated such that children come to have correct expectations about when others will perform actions which are contributions to outcomes which they desire or toward which they are acting. Similarly, motivations are calibrated such

that children come to be motivated to make contributions when they are expected to -- and particularly when it is important to others that they do so, and particularly when the other person in question is one with whom it is important to maintain a good relationship. The upshot of this account is that proficiency in generating commitments, and in identifying and tracking the degree of one's own and others' commitments crucially involves managing expectations about contributions to goals and desired outcomes.

Where does all this leave the concept of commitment and the simple conjecture? Mastering the concept of commitment in the strict sense does not appear to be necessary in order to identify and respond to such expectations on the part of others, or to have such expectations about others. Nor does it appear to be sufficient in order to (a) determine when commitments are in place in the absence of an explicit agreement or promise, to (b) determine what the precise content of an explicit or implicit commitment is, to (c) assess the appropriate degree of commitment, or (d) to distinguish between good and bad reasons for abandoning commitments. However, this does not make the concept of commitment in the strict sense irrelevant. On the contrary, there are several important functions made possible or facilitated by master of the concept of commitment in the strict sense.

For example, mastery of this concept makes it possible to quickly and efficiently engage the machinery of expectations and motivations that we have been attempting to illuminate here. Doing this proficiently, however, also requires that one's expectations and motivations are properly calibrated to begin with. For example, if Orsi gives Vanda an assurance that she will clean up every mote of dust that ever falls onto his car, he is unlikely to form the expectation that she will actually do this, because it is simply not a realistic suggestion. Similarly, if she requests after their first date that Vanda promise to be forever true, it might well have the opposite effect, because it is an unreasonable request, and indeed one which exhibits an alarming lack of social skill.

Moreover, the concept may help in various ways to facilitate the calibration of motivations and expectations that we have been discussing. For example, the concept of commitment in the strict sense highlights some features of situations that are relevant to determining whether ME can reasonably expect YOU to do X, such as whether YOU did something to generate this expectation in ME, whether this was intentional, and whether it is common knowledge that this is the case. These are not the only relevant factors, but they are among the relevant factors. So if Daddy promises to give Leonardo some ice cream after dinner and then only gives him a single scoop, and Leonardo begins to cry and protest about this, Daddy may point out to him that he promised to give him only a bit and was never

intending to suggest that he would give him any more -- Leonardo will have to calibrate his expectations downward about what 'some ice cream' means.

5. Conclusions

We humans quite proficient in generating commitments, and in identifying, keeping track of, and responding appropriately to our own and others' commitments. In the current paper, we have attempted to shed light on the cognitive processes underpinning this proficiency, in particular by examining the emergence of a proficiency in managing commitments in ontogeny.

One unsurprising general conclusion to draw is that humans, armed with the concept of commitment and with the language skills to make verbal agreements and otherwise to form and communicate detailed plans about future behavior, are highly adept at generating expectations, which others can rely on. It would also be unsurprising if some of the source of the motivation to fulfil those expectations are uniquely human.

One perhaps surprising consequence of our account is that a very powerful source of motivation to fulfil those expectations, and basis for expecting others to do so as well, is in fact the product of a very basic tendency that is present early in ontogeny and likely shared with other species -- namely, a tendency to become frustrated and angry if our goals are not met and the outcomes we desire not achieved. Specifically, our account generates a novel claim about the origins of the sense of entitlement that inspires protest over unfulfilled expectations, i.e. unfulfilled expectations about one's goals being met and about the outcomes one desires coming about. In contrast to the hypothesis suggested by the simple conjecture, our account suggests that this sense of entitlement to protest is not the product of developmental changes by which one acquires the concept of commitment but, rather, it is the default that is already in place by two or earlier. What changes over the course of childhood is that children learn that they are not always entitled to expect the goals to be met or all contributions to their goals to be made. In other words, the developmental process chips away from, rather than adding to, the cognitive architecture that underlies normative protest.

Author Biographies

John Michael completed his PhD in philosophy at the University of Vienna in 2010. Since 2016 he has been Assistant Professor of Philosophy at the University of Warwick and Affiliated Researcher the Department of Cognitive Science of the Central European University in Budapest. His research interests include the sense of commitment, self-control, joint action, perspective-taking and other issues at the intersection between philosophy and cognitive science. He currently holds an ERC starting grant investigating the sense of commitment in joint action.

Marcell Székely studied medicine at the Semmelweis University, and psychology at the University of Glasgow and Plázmány Péter University. He is currently a research assistant at the Central European University. His research interests include commitment, effort, motivation perspective-taking, and other issues in social cognition and behavioral economics.

Acknowledgments

This research was supported by a Starting Grant from the European Research Council (n 679092, SENSE OF COMMITMENT) and by the European Project CODEFROR (FP7-PIRSES-2013-612555).

References

Aarts H, Dijksterhuis AP (2000) The automatic activation of goal- directed behaviour: the case of travel habit. *J Environ Psychol 20(1):75–82*

Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie my hands, tie my eyes. *Journal of Experimental Psychology: Human Perception and Performance, 38(2), 263.*

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

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental psychology*, 41(2), 328.

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

Brownell, C., G. Ramani, and Zerwas, S. (2006). Becoming a social partner with peers: cooperation and social understanding in one- and two-year-olds. *Child Development*, 77 (4), 803-821.

Butterfill, S. (2012). Joint action and development. *Philosophical Quarterly,* 62 (246), 23-47.

Butterfill, S. and Apperly, I. (2013). How to construct a minimal theory of mind. *Mind and Language*, 28 (5), 606-637.

Butterfill, S. A., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, *88*(1), 119-145.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*, 159-181.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255-259.

Fiebich, A., & Coltheart, M. (2015). Various ways to understand other minds: Towards a pluralistic approach to the explanation of social understanding. *Mind & Language*, *30*(3), 235-258.

Gilbert, M. (1990). Walking together: a paradigmatic social phenomenon *Midwest Studies in Philosophy*, 15, 1-14.

Gilbert, M. (1989). *On Social Facts*. London: Routledge and Kegan Paul.

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

Gilbert, M. (2006b) *A Theory of Political Obligation*. Oxford: OUP.

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

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, 45(5), 1430-1443.

Gräfenhain, M., Carpenter, M., Tomasello, M. (2013). Three-Year-Olds' Understanding of the Consequences of Joint Commitments. *Public Library of Science ONE* 8(9): e73039. doi:10.1371/journal.pone.0073039

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

Heintz, C., Celse, J., Giardini, F., & Data, S. M. (2015). Facing expectations: Those that we prefer to fulfil and those that we disregard. *Judgment and Decision Making*, *10*(5), 442.

Jacob P, Jeannerod M (2005) The motor theory of social cognition: a critique. Trends Cogn Sci 9(1):21–25

Mant, C. M., & Perner, J. (1988). The child's understanding of commitment. *Developmental Psychology*, 24(3), 343-351.

Costantini, M., Ambrosini, E., Cardellicchio, P., & Sinigaglia, C. (2013). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience*, *9*(5), 705-711.

Michael, J., & Székely, M. (2017). Goal slippage: a mechanism for spontaneous instrumental helping in infancy?. *Topoi*, 1-11.

Michael J, Christensen W (2016) Flexible goal attribution in early mindreading. Psychol Rev 123(2):219–227

Michael J, Sebanz N, Knoblich K (2016) The sense of commit- ment: A minimal approach. Front Psychol 6:1968. doi:10.3389/ fpsyg.2015.01968

Michael, J. & Pacherie, E. (2015). On Commitments and Other Uncertainty Reduction Tools in Joint Action. *Journal of Social Ontology*,1(1): 89-120.

Piaget, J. (1950). *The psychology of intelligence*. New York: Harcourt, Brace.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental psychology*, *44*(3), 875.

Scanlon, T. (1998). *What we owe to each other*. Cambridge: Harvard University Press.

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.

Searle, J. (1969). *Speech Acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

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

Sinigaglia, Corrado, and Stephen A. Butter ll. 2016. 'Motor Representation in Goal Ascription'. In Foundations of Embodied Cognition 2: Conceptual and Interactive Embodiment, edited by Yann Coello and Martin H. Fischer, 149–164. Hove: Psychology Press.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-B11.

Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: the impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental psychology*, *44*(5), 1249.

Tollefsen, D. (2005). Let's pretend: children and joint action. *Philosophy of the Social Sciences,* 35 (75), 74-97.

Tomasello, M. (2009). *Why we cooperate*, Cambridge: MIT Press.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124-130.

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

Warneken F, Tomasello M (2007). Helping and cooperation at 14 months of age. *Infancy,* 11, 271–294. doi: 10.1111/j.1532-7078.2007.tb00227.x