

The Social Motivation Hypothesis for Prosocial Behavior

Philosophy of the Social Sciences
2014, Vol. 44(5) 563–587

© The Author(s) 2014

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0048393114530841

pos.sagepub.com



Marion Godman^{1,2}, Michiru Nagatsu¹,
and Mikko Salmela¹

Abstract

Existing economic models of prosociality have been rather silent in terms of *proximate* psychological mechanisms. We nevertheless identify the psychologically most informed accounts and offer a critical discussion of their hypotheses for the proximate psychological explanations. Based on convergent evidence from several fields of research, we argue that there nevertheless is a more plausible alternative proximate account available: the social motivation hypothesis. The hypothesis represents a more basic explanation of the appeal of prosocial behavior, which is in terms of anticipated social rewards. We also argue in favor of our own social motivation hypothesis over Robert Sugden's fellow-feeling account (due originally to Adam Smith). We suggest that social motivation not only stands as a proximate account in its own right but also provides a plausible scaffold for other more sophisticated motivations (e.g., fellow-feelings). We conclude by discussing some possible implications of the social motivation hypothesis on existing modeling practice.

Received 17 March 2014

¹University of Helsinki, Helsinki, Finland

²University of Cambridge, Cambridge, UK

Corresponding Author:

Michiru Nagatsu, Social and Moral Philosophy/Department of Political and Economic Studies,
University of Helsinki, PL 24 (Unioninkatu 40 A), 00014 Helsinki, Finland.

Email: michiru.nagatsu@helsinki.fi

Keywords

social motivation, pro-social behavior, fellow-feeling, social norms, team reasoning

I. Introduction

It is widely accepted that a distinctive aspect of human sociality is the pervasive engagement in “prosocial behavior,” which can be roughly characterized here as an individual’s choice to forgo her immediate material interests to benefit other(s) or a group to which she belongs. Many behavioral economic models hypothesize some proximate motivations in addition to purely material self-interest to explain such prosocial behavior. However, the proposed psychological mechanisms are often rather casually proposed to accommodate behavioral data in stock experimental games and are often poorly informed by psychological research. Moreover, contrary to the stated ambitions, this strategy so far remains unsuccessful in delivering a unified picture of what motivates prosocial behavior (Clavien and Klein 2010).

This article was triggered by the following observation: perhaps there are too many models and not enough attention to the empirical work in various areas of psychology. It seems as though we should try to get the psychological (proximate) mechanisms right before modeling a particular motivation as an augment of an individual utility function, or presuming that such modeling strategy is indeed the best way to proceed. We support this move while acknowledging that like any empirical hypothesis, psychological ones are fallible and in principle open to revision. Still, we hope that this article will not only provide the reader with a psychologically realistic and plausible proximate account of prosocial behavior but also prompt some general methodological reflections on how psychological research bears on existing modeling practice.

We proceed as follows. In section 2, we consider the two psychologically most informed economic modeling strategies for explaining prosociality and argue that they are, at least as it stands, unsatisfactory with respect to providing proximate accounts of prosocial behavior. In section 3, we consider Adam Smith’s fellow-feeling hypothesis revived by Robert Sugden (2002), as a more promising alternative for the motivations for norm-following/team-reasoning. In section 4, we present a different social motivation hypothesis as well as some convergent evidence for it, and in section 5, we argue that social motivations might be understood as a scaffold for the more specific rewards involved in fellow-feelings and, analogously, as a general scaffold for other candidate proximate accounts of prosocial behavior. Section 6 concludes with some methodological reflections.

2. Economic Models of Prosocial Behavior

Prosocial behavior, such as cooperation in the prisoner's dilemma (PD) and public goods games (PG), fair offers in ultimatum bargaining games (UG), and trusting and sharing in trust games, is a fairly robust experimental phenomenon—even in those experiments where one is careful to ensure that the subjects believe that interactions are anonymous and that there is no future interaction with other subjects (see, for example, Camerer 2003, chapter 2 for a review).

We can identify at least four approaches within the rational choice tradition that purport to explain such experimental results as well as our prosocial behavior in analogous real-life situations. The first two are (i) the rational cooperation in repeated games approach, according to which people mistakenly behave prosocially in one-shot games because humans have evolved in repeated interactions where cooperation is rational (e.g., Binmore 2006); and (ii) the social preference approach, which models people's motivations as stable "social preferences" for equity, fairness, and so on that are defined in terms of the material payoffs of games (e.g., Fehr and Schmidt 1999). Although popular among behavioral economists, there is convincing empirical evidence against these hypotheses (see, for example, Gächter and Thöni 2007 for a review). Moreover, these accounts do not explicitly address motivations of prosocial behavior: the first (i) simply assumes such motivations to be explanatorily irrelevant and the second (ii) crudely equates prosocial motivations with preferences for different distribution of players' material payoffs in games. We, therefore, focus on the two other approaches in this paper: (1) social norms and (2) team-reasoning.

Both (1) social norms and (2) team-reasoning seem to hold the most promise for explaining a range of experimental results (see Tan and Zizzo 2008; Guala, Mittone, and Ploner 2013, for empirical evaluations). The approaches are also (though not always) accompanied by explicit discussions of what motivates norm adherence and team-reasoning, respectively. However, after a discussion of each approach, we will briefly argue that these proximate accounts are ultimately unsatisfactory.

2.1. Social Norms Approach

The *social norms approach* has been gaining popularity due to the growing evidence that specific expectations (beliefs) as well as preferences influence prosocial behavior in economic games. Unlike the social preferences approach, which assumes some stable other-regarding preferences at work in all contexts, the social norm approach adopts the idea of *conditional*

preference. One of the main proponents of the approach, Christina Bicchieri thus argues that the preference to follow a norm or behavioral rule is activated only if the individual believes (a) that enough others will follow the rule (“empirical expectation”) and (b) that enough others will expect her to follow (“normative expectation”) (Bicchieri 2006). Due to this conditionalization of preferences as relative to certain expectations and perhaps also what norms are identified, the social norms have a positive explanation for why we should expect (some) prosocial behavior to occur in some contexts (e.g., ultimatum games with humans) and not in others (e.g., ultimatum games with computers) (see Bicchieri 2006, 125, also 168-170). Once these epistemic conditions of norm identification are met, Bicchieri’s model predicts that players’ payoffs in a game are transformed. That is, players’ payoffs for different outcomes of the game are subjectively modified to take into account players’ taste for certain social norms (of fairness, cooperation, etc.). This changes the payoff structure of the PD and other economic games by, for example, rendering defection or “unfair” behavior less attractive, which in turn makes their overall self-interest aligned with prosocial norms. Social norms “tug at our utility” of what is economically rational (Paternotte and Grose 2012). The approach thus models individual utilities such that they involve preferences for norm conformity.

But what precisely motivates people to follow social norms, once the epistemic conditions of norm identification are met? Insofar as the social norm approach can go beyond just describing the *behavior itself* in terms of conditional preferences or stipulate that the mere identification of the norm itself automatically triggers norm-following, the model owes us a *proximate* explanation for *why agents are motivated to follow the norm and adhere to others’ expectations*.¹ In other words, the mere activation of a social norm (or an expectation that there is such a norm at work) is not sufficient since there must be something that explains why empirical and normative expectations *matter* to individuals.

The same question is provoked by Herbert Gintis’s (2010) model of social norms, which, just like Bicchieri’s, presupposes that people have preferences

¹Although Bicchieri (2006, 3) explicitly states that her belief-desire model of choice is compatible with the possibility that our norm-following behavior occurs subconsciously without deliberate decision, she does not understand this folk psychological model as an “as if” model.

to follow social norms.² Gintis (2011), however, stresses that certain *moral values* and *character virtues* lead individuals to voluntarily conform to norms. These values and virtues include promise-keeping, fairness, and honesty, and they manifest as behavioral regularities that facilitate cooperation and enhance social efficiency (Gintis 2011, 884). According to Gintis, individuals internalize these behavioral regularities through processes of socialization where the moral values and character are instilled by the interplay of affect and authority and induce us to comply and conform to the specified norms that promote prosocial behavior. This internalization means agents behave prosocially even if there is no possibility of being penalized for selfish behavior and being rewarded for prosocial behavior. “Morality, in this way, is doing the right thing even if no one is looking” (Gintis 2010, 252).

Now, Gintis admits that moral character and virtues do not always guarantee norm abidance as other conflicting interests are also present. Yet insofar as we act prosocially, Gintis believes that the central motivation for such behavior is normative or moral: doing the right thing in the situation. Still, it seems that Gintis might be too optimistic in thinking that moral and character virtues alone are responsible for inducing norm abidance. Recent empirical work in moral psychology shows that those moral principles and character, including virtues, which are used to justify actions, do not appear to have any “formative” role in our decision-making (for a review, see Merritt, Doris, and Harman 2010). Some even suggest that moral reasoning is a result of confabulation (Schnall, Benton, and Harvey 2008; Wheatley and Haidt 2005). Either way, the role of moral traits and virtues as proximate explanations has not been confirmed to the extent assumed by Gintis.³

Bicchieri (2006, 13-25) provides a more nuanced and extended proximate account. She suggests three types of motivations for norm adherence: (i) fear of resentment (and of more consequential punishment), (ii) desire to please others’ expectations and preferences, and (iii) belief that others’ normative expectation is legitimate. Bicchieri places most credence in (iii) and argues that acceptance of certain norms (e.g., cooperation in PG or fair offer in UG) as *reasonable* or *legitimate* is the more important proximate explanation. In interactions with family and friends, a generic desire to meet others’

²Gintis (2010) models social norms as a coordination device (the “choreographer”) that enables players to choose one from multiple Nash equilibria, using the concept of correlated equilibrium. See for further comparisons of Bicchieri’s and Gintis’s models.

³In addition, Gintis does not specify the epistemic conditions under which once-internalized norms are triggered (see Paternotte and Grose 2012).

expectations will suffice, but in most other interactions with socially distant others, she contends that generic desires (i) and (ii) have little motivating power, and so the beliefs about social norms being legitimate are required.

It is undeniable that people are sometimes motivated to meet others' expectations because they find those expectations reasonable or legitimate. But it seems that norm conformity is not always associated with the perception of reasonableness. For example, Solomon Asch's (1955) classic experiments show that an individual's easy judgment about the length of a line was significantly influenced by the answer given by other members of a group when the answer was obviously wrong but unanimous. In such cases, conformity alone seems to motivate even if the expectations of others are unreasonable. Thus, conformity seems to be de-coupled from the perceived reasonableness of the norms (see Lisciandra, Postma-Nilsenová, and Colombo 2013 for a similar conformity effect on normative judgments). In other words, people seem to behave as if they just want to conform to the majority's opinions and preferences, regardless of whether it is reasonable to do so or not.

In fact, once we turn to Bicchieri's own model, motivations of types (i) and (ii) seem required to derive (iii). In her account of conditional preferences, Bicchieri (2006, 15) distinguishes two types of "normative expectation," that is, one's belief that a sufficiently large subset of population expects one to conform to a behavioral regularity in a type of situation. The first type is an *empirical expectation* (i.e., belief that others think one *will* conform), and the second, *normative* (i.e., belief that others think one *ought to* conform).⁴ What motivates us to act according to such expectations? It seems that viewing such expectations as legitimate or reasonable is not going to do the work in either case: the expectation that others think one will conform does not in and of itself make the expectation legitimate—it is just a majority opinion about what will likely be the case. Similarly, a normative belief (that others think one ought to conform) will not be viewed as reasonable unless one thinks that the majority opinion itself is a source of legitimacy. In both cases, legitimacy or reasonableness in Bicchieri's sense seems to be ultimately grounded in the majority opinion and the majority opinion alone can motivate only if one already has a conformist preference of types (i) and (ii).

⁴So Bicchieri's terminology is slightly unfortunate. When she calls two epistemic conditions "empirical" and "normative" expectations, she only means that the former is a belief about others' behavior, while the latter is a belief about others' beliefs about one's behavior.

For this reason, her distinction between (i) and (ii) as a generic conformist desire and (iii) as reason, as well as her emphasis on the latter, does not hold.⁵

In contrast to Gintis and Bicchieri, Robert Sugden (2000) models social norms explicitly based on the *non-normative* conformist desire. He introduces a *resentment hypothesis*, according to which we are motivated to meet the expectations of others because we are averse to their resentment—a resentment that we expect will be provoked by frustrating their expectations. Sugden takes this to be an empirically confirmable fact about human psychology; that is, people resent someone who violates their well-established empirical expectations, and people are averse to face such resentment themselves. Supposing we do find others' expectations "legitimate," then this will be explained by wanting to please them because not doing so will provoke their resentment. So to use Bicchieri's distinctions, the type (iii) motivation (to do something legitimate) is grounded in type (ii) motivation (to please others' expectations), which in turn derives from the more basic type (i) motivation (to avoid others' resentment).

We agree that the aversion toward being a target of resentment might motivate the conditional preferences to meet others' expectations in Bicchieri's model.⁶ Still, it is questionable whether aversion to of resentment is the type of psychological feature that can explain a wide range of prosocial behavior, including cooperation in one-shot PG games where one has no reason to expect cooperation as a norm. Moreover, it raises the question of why others' resentment should matter to us anyway—especially if they are strangers. All the same, we do not wish to argue against the resentment hypothesis at this stage. We believe that the resentment hypothesis has a legitimate place as a proximate account but only once one finds an appropriate mechanism that explains why resentment should matter to us in the first place—a task we will turn to in the succeeding sections of this article. In section 3, we will see how Sugden himself suggests what might possibly scaffold such resentment, but in section 4 and onward, we will argue that the social motivation hypothesis can provide a superior and more basic account of the resentment hypothesis.

⁵In Bicchieri's examples of social norm engineering, such as anti-littering and a campaign against college students' binge drinking, people's behavior changes not by changing their belief that a certain behavior is legitimate or not but by changing the perception of how widespread a certain behavior is. Such an intervention also presupposes a sort of generic preference for conformity.

⁶Sugden (2000) sketches his own model of social norms, but to our knowledge, no one, including himself, has tested this model.

2.2. Team-Reasoning Approach

A prominent contender of the social norm approach to explaining prosocial behavior is the team-reasoning approach (Bacharach 2006; Coleman, Pulford, and Rose 2008; Sugden 1993, 2000). Here, the prosocial behavior is modeled by an individual *identifying herself as a member of a team, and thus playing her part to achieve the group's objective*. If a player identifies herself as a member of a group consisting of all or most of the players, then it is rational action to cooperate or coordinate with others, since it will uniquely maximize the group's collective payoff. The account then involves an *agency transformation* that is distinct from the payoff transformation of the social norm approach.⁷

The team-reasoning approach needs to specify the epistemic conditions for the presumed agency transformation—or shift from I-mode to we-mode, to use Tuomela's (2007) terminology. Bacharach (2006), who offers one of the most systematic presentations of the team-reasoning approach to date, proposes the “interdependence hypothesis” that roughly states that the agency transformation is triggered if the agents perceive common goals that can only be achieved together (for conditions for agency transformation, see Bacharach 2006, 83-85; Smerilli 2012). The epistemic conditions of the interdependence hypothesis are thus provided by such features representing the uncertain prospect of mutual benefit.

For Bacharach, the shift in framing occurs spontaneously and is not a matter of choice. Nevertheless, even supposing that conscious choice typically does not enter into the agency transformation, the account still seems to require some proximate account of the psychological features in the individual agent that are responsible for such a shift (Pacherie 2011). For unlike a purely perspectival shift between alternative ways of looking at a *gestalt* figure, the framing effect relevant for agency transformation does seem to require some sustained proximate account with both cognitive and motivational elements. Consider, for example, the analogous reference-dependent

⁷Often, the team-reasoning model is misunderstood as just another other-regarding preference model where “other” regards the group to which an individual belongs. An analysis of the Hi-Lo game shows this is not so. A Hi-Lo game is a two-person, two-by-two, normal form, symmetric pure coordination game in which one Nash equilibrium (Hi, Hi), say (2, 2) pareto-dominates the other (Lo, Lo), say (1, 1). The individuals' payoff transformation that takes into account the group's payoff changes the payoff of (Hi, Hi) as (4, 4), (Lo, Lo) as (2, 2), which leaves both as Nash equilibria. In contrast, the agency transformation makes (Hi, Hi) a unique equilibrium of the game.

framing effect (Tversky and Kahneman 1981), which is purportedly proximally explained by loss-aversion.

While Bacharach (2006, chapter 3) suggests that group selection can account for why propensities to team reason have arisen under conditions consistent with the interdependence hypothesis, he does not give any detail as to what psychological propensities account for the identification of such features. Indeed, it is not only team-reasoning accounts that are typically silent on what motivates the initial and continued attachment to the group. This problem is common with an influential philosophical theory of collective intentionality (Gilbert 2013), which presupposes a *joint commitment* of the members to their group's goals but does not give an account of proximate motivation for such commitment (Pacherie 2011).

It is notable that the resentment hypothesis, which was intended as a complement to the social norm approach, is not applicable in the case of team reasoning. Since a key epistemic condition spelled out in the interdependence hypothesis is a perception of a mutually beneficial but *uncertain* outcome, this seems to block the possibility that it is the fear of resentment that motivates agent transformation (since why would such circumstances provoke resentment?). Moreover, presumably collective intentionality theorists would also hesitate to posit fear of resentment as a motivational basis of a joint commitment, since that would reduce the model of collective intentionality to one of individual norm-following like Bicchieri's.

2.3. Summary

In this section, we have discussed the proximate hypotheses of the most psychologically informed economic approaches to prosocial behavior: the social norm and team-reasoning approaches. Our first result is that Bicchieri's model of social norms presupposes some motivational account and that Sugden's resentment hypothesis is currently the best candidate. However, we also suggested that the hypothesis seems to call for an explanation of why the resentment of others should matter to us and noted that it is illfitted to deal with the team-reasoning approach (which lacks any explicit proximate account of its own). We now move on to some general proximate accounts that are more promising but will have reasons to return some of the hypotheses of this past section (like the resentment hypothesis) in section 5 to show how they can be scaffolded on more basic social motivations.

3. The “Fellow-Feeling” Hypothesis

In his 2002 article, Sugden admits that team-reasoning lacks a motivational dimension just like other rational choice models. His remedy draws heavily

on Adam Smith's ([1759] 2002, 82) work *Theory of Moral Sentiments* and suggests that Smith's account of *fellow-feeling* "may be able to fill these gaps by explaining the affective qualities of team thinking." While Sugden clearly takes the approach as a complement to the team-reasoning account, rather than the social norm account, we will consider the fellow-feeling hypothesis as a potential psychological account of prosocial motivation in general, that is, for both norm-following and team-reasoning.

According to Sugden, Adam Smith suggests that simply doing things together with others rather than alone creates added value to activities by giving a certain satisfaction or reward. This reward is due to participants' positive affective awareness of their corresponding sentiments. This is also the definition of fellow-feeling: *the pleasant awareness of corresponding sentiments or emotions*. Smith claims that this fellow-feeling is rewarding irrespective of the valence of original sentiments. Thus, fellow-feeling cannot be reduced to mere reflected emotions: "this correspondence of sentiments of others with our own appears to be a cause of pleasure, and want of it a cause of pain, which cannot be accounted for 'by a theory of reflected feelings'" (Sugden 2002, 72). The *agreement of feelings*, be they joy or grief, is thus a source of reward or pleasure in its own right such that fellow-feeling "enlivens joy and alleviates grief" (Smith [1759] 2002, 18).

According to Smith, our desire to experience fellow-feelings motivates us to align our actions and emotions with those of others. Moreover, those that have a history of repeated interactions will be more likely to experience such fellow-feelings:

On Smith's account, it is a fact of human psychology that people who repeatedly interact with one another tend to develop and express common sentiments. Such common sentiments tend to eventually become the objects of common approval within the group of interacting people. Thus, the observed failure of any one member of a social group to uphold the attitudes of that group will cause pain or unease to other members (this is just the negative equivalent of the pleasure of mutual sympathy); and it will be disapproved of. (Sugden 2002, 82-83)

The account can thus be seen as a proximate complement to the existing prosocial models since people become motivated to adjust their emotions and behavior to fit with the communal "norms of propriety" or else they will forgo opportunities for fellow-feelings. That is, the pleasure derived from fellow-feelings motivates individuals to uphold certain shared emotions and behavioral regularities that are associated with them. The pleasure associated with fellow-feelings also motivates individuals to identify with groups that

they have interacted with in the past and to uphold the communal norms of these groups. Alternatively, as Bicchieri would put it, fellow-feelings give individuals a reason to comply with *empirical* and *normative* expectations. While a generic desire to meet other people's expectations cannot explain our context- and group-specific prosocial behavior (Bicchieri 2006, 25), the fellow-feeling hypothesis implies that we are only motivated by those expectations that have arisen during those previous social interactions that included shared sentiments. True fellow-feeling *is* a generic affect for both Sugden and Smith, but repeated interactions with fellow-feelings are needed for the generic emotion to reliably motivate people to follow context- and group-specific social norms.

The affective rewards that the individuals derive from the interaction and the motivations that precede them are thus valuable in their own right. In Sugden's (2002, 85) own words, "we *directly* desire to participate in society, and not merely the ends that it produces." Nevertheless, the motivational power of the fellow-feeling proposal relies on rewards that are only attainable from shared emotions in social interaction and dependent on awareness of other participants' corresponding emotions. In the next section, we will argue that because of this the fellow-feeling approach is still too restrictive and that correspondence or mutuality of sentiment might not be a prerequisite for all experiences of intrinsic social rewards. The model we present will thus be a weaker and also a more basic model of social motivation. The social motivation here derives from social-affiliative stimuli that does not demand aligned emotions.

4. The Social Motivation Hypothesis

The pleasurable reward in the fellow-feeling account was connected to a distinctive type of social encounter. In such encounters, emotions of the participants are aligned such that all may reap enjoyment from it; but why suppose that it is only on these occasions that sociality is rewarding? There seems to be a broader array of affiliative stimuli, that is, interactions with, or observations of, other humans, which have the potential to generate intrinsically rewarding experiences. These occasions are often consciously recorded as warmth, affection, and other positive affect connected to affiliation.⁸

⁸In the theory of interpersonal behavior, *affiliation* is one of the two major independent traits characterized by warmth, positive emotions, and agreeableness (the other is *agency* characterized by assertiveness and activity). To represent social stimuli that are affiliative, experiments typically use pictures of smiling babies and families (see Depue and Morrone-Strupinsky 2005, 315).

Neuroscience also suggests that the set of basic biological and psychological mechanisms that generate these rewards are elicited by the same modalities of sensory affiliative stimuli (i.e., touch), even if the social activities that generate the sensory stimulation are different (Crews 1998; Depue and Morrone-Strupinsky 2005; Mason and Mendoza 1998). Although activities involving *shared* emotional experiences certainly induce these rewards, several other activities are also capable of this effect, for example, those that are more purely coordinative, explorative, and even more competitive in character (think of how academics enjoy a verbal spar even if this activity often involves antagonistic emotions between the disputants).⁹ Sometimes the reward may even be elicited by something as fleeting as someone smiling at you.

It is commonly thought that the role of the neurobiological substrates underlying the affective rewards of affiliative stimuli (oxytocin, vasopressin, endogenous opioids) is to orient and bias attention to very basic forms of mammalian affiliation (parent and child) as well as forming social relationships more generally (Depue and Morrone-Strupinsky 2005; Nelson and Panksepp 1998). The experience of rewards attached to social interaction and affiliation generates social motivation circuitry that not only motivates us to seek the pleasure in social interactions (affective reward) but also works to establish and facilitate social bonds (Godman 2013). Many social bonds in other primates are also thought to be mediated or enhanced by rewards elicited during activities such as play and grooming (Silk et al. 2009). The social rewards elicited by affiliative stimuli thus provide a foundation for the capacity for forming social bonds crucial to human development and flourishing (see, for example, Baumeister and Leary 1995).

So how precisely does the experience of a reward attached to social-affiliative stimuli lead to a social motivation? The thriving general field of motivation research backs up the Humean assumption about how motivations typically grow out from an anticipation of a reward (Berridge 2003; Wise 2004). In the case of social affect, the idea would be that affiliative stimuli

⁹Academics enjoy the spar and thus get reward from it, but their enjoyment is not shared in a robust sense (see Salmela 2012). Their emotion is shared only in a weak sense of being the same type (enjoyment) and having the same intentional object (the spar). Even if the participants enjoy the spar, their concerns are directly opposite within this activity: both strive to win but since spar is a zero-sum game where one agent's win is another's loss, their emotions do not correspond within the activity. Accordingly, no fellow-feelings emerge.

initially become “incentively salient” due to the biased capacity for experiencing such rewards. The initial link between affiliative stimuli and social reward is then strengthened in terms of basic associative learning. That is, because rewards are typically attached to certain affiliative stimuli, they become increasingly associated in such a way that they yield (social) motivations in expectation of distinctively social rewards. Thus, due to the intrinsic rewards afforded by affiliation and sociality, we become ever more oriented toward the social world. But just like in the case of other rewards, sometimes, wanting eventually occurs without the accompanying conscious, or unconscious, experience of “liking,” thereby indicating how motivations can be de-coupled from the associated reward (Berridge and Robinson 2003). (This may indeed be the case in many economic games where subjects are naturally conditioned to a particular social reward that is not in fact elicited).

The social motivation hypothesis then contends that there is a particular psychological disposition whose role is to orient us toward affiliative stimuli, which yields social reward (affect) and enables the formation of social bonds. In fact, the hypothesis is nicely captured by Coralie Chevallier and colleagues who defend a social motivation hypothesis in a rather different context, namely, in the case of autism (more on this in a moment):

Social motivation can be described as a set of psychological dispositions and biological mechanisms biasing the individual to preferentially orient to the social world (social orienting), to seek and take pleasure in social interactions (social reward), and to work to foster and maintain social bonds (social maintaining). (2012, p. 231)

It is important to emphasize that the kind of affiliative stimuli that people seek and find rewarding depends crucially on their past experiences and learned associations. Some individuals may actually become motivated to seek antagonistic social encounters (e.g., winning academic debates or fighting with partners), perhaps because they have found other rewards in such activities or even because antagonism is actually a condition of affiliating with others. On the other hand, if one constantly meets rejection and ostracism in trying to affiliate with particular others, typically one will feel bad and eventually give up, or try to find social rewards elsewhere. In short, social motivations are contingent upon their history of particular learned associations and although at a baseline we take some initial affiliative stimuli to be rewarding in their own right, we are not committed to claiming that all social stimuli must remain rewarding and sought for.

In recent years, there has been several studies in different domains of developmental psychology and psychiatry that have suggested that social

motivation arises early in development and is a crucial feature of human psychology. Social motivations have been suggested to not only mediate and facilitate bonding but also to explain autism and the pervasive tendency for social learning (i.e., learning by copying others, rather than trial-and-error learning). Here, we will only mention a couple of examples of such convergent evidence.

First, it is well-established that infants are born with a preferential ability to track and fixate stimuli that appear human-like, and evidence indicates that this is mediated by the neurological center for affective processing—the amygdala (Pierce et al. 2001; Young et al. 1995). In addition, recent research shows how the amygdala is involved in quickly detecting salient and diagnostically relevant facial features by enabling attentional orienting toward the visual periphery (Gamer et al. 2013). These results suggest that we are preferentially oriented toward one of the most basic form of social-affiliative stimuli—the facial expressions of others—due to the associated affective reward.

Recently, social motivation has also become an important rival hypothesis to the standard “theory of mind deficit” approach to explaining autism (Chevallier et al. 2012). The model suggests that autistic subjects’ difficulties with social life are better accounted for by diminished social motivation rather than a failure of “theory of mind,” precisely because the perception of social-affiliative stimuli, such as facial expressions, gestures, and vocalizations, are not accompanied by any distinctive type of reward. The lack of social motivation, thus, impedes social development and can account for many of the interpersonal deficits associated with the autistic spectrum disorder (Chevallier et al. 2012; Maestro et al. 2002).

Finally, social motivations have been suggested to play a prominent role in human ontogeny in biasing social learning over individual learning. Cecilia Heyes (2012, 7) has, for example, suggested that what explains abundant human social learning strategy is social motivations and other perceptual biases that privilege information received via social channels over any individual learning channels. Some evidence for this comes from work on imitation where social motivations are a strong contender for explaining the pancultural human propensity for so-called *overimitation*, that is, the tendency to reproduce a high-fidelity match of the model’s behavior to the extent that the efficiency of the task is reduced (Nielsen and Tomaselli 2010). In cases of overimitation, we do *not* reach the goal in the most effective way possible but try to do it in precisely the way it has been modeled to us. For example, children begin to overimitate at the age of two but by no means cease to do so when they become adults (McGuigan, Makinson, and Whiten 2011). Social motivations provide a compelling explanation for why this is

so. That is, reproducing the modeled actions appears to be more important than merely producing the outcomes more efficiently, precisely because we are socially motivated to engage in the activity. (Over)imitation is not merely a means of learning skills or acquiring information about the world; it is also a means of engaging with others and attaining social rewards (Nielsen and Blank 2011; Uzgiris 1981). The social motivation thesis thus predicts that whenever the opportunity for social rewards is a particularly salient part of the target for learning, *how* the effect is produced will matter more than that it is produced in the most effective way possible.

Taken together this evidence from different strands of psychology also suggests an evolutionary argument for the emergence of social motivations. Most likely, social motivations originated from their role of facilitating and maintaining social bonds—a role that renders them not exclusively human, but among humans, it is likely that they are a particularly strong motivator (e.g., non-human animals do not tend to overimitate). Indeed, it may be part of the explanation for why humans are regarded as the “hyper-social species.” Of course, many tie the trait of hyper-sociality to the emergence of culture, social norms, and large-scale cooperation (Boyd and Richerson 2009), and at the very least once culture enters into the picture, it is highly likely that culture will co-opt some of the psychological bias for social reward into new roles. One such is their role in facilitating social learning and the fidelity of social transmission; another is further biasing individuals to prosocial behavior. Indeed, it is natural to suppose that social motivations are precisely the sort of psychological capacity that can be harnessed to achieve common, indeed, genuinely altruistic ends that involve overall material sacrifice for the individual, such as when social motivations help to facilitate cooperation and cohesion among groups.¹⁰

5. Social Motivations as Scaffolders

The social motivation hypothesis is more general than the fellow-feeling approach in terms of what type of stimuli yield an affective social reward. For affiliative stimuli to be rewarding, there is no requirement of fellow-feeling, or for that matter shared emotions in social activities. In this section, we will

¹⁰While some cooperative endeavors might only be done by groups, or are more efficiently performed by groups, and so are not genuinely altruistic as they are aligned with self-interest; others might demand foregoing material self-interests. Researchers disagree on which mechanism (self-sacrificing or self-interested prosociality) has more explanatory relevance (see Guala 2012).

consider the relationship between the two accounts more closely and review some evidence from synchrony studies that offers support to our more general social motivation hypothesis over the fellow-feeling account. However, our more important aim is to argue that social motivations should be thought of as a scaffold for the emergence of fellow-feelings and, analogously, for other proximal explanations of prosocial behavior.¹¹

Let us first review some of the evidence from the synchrony studies that counts in favor of the social motivation over the fellow-feeling hypothesis. The association between synchronized behavior, on the one hand, and affective reward, on the other, has been established in several different studies (e.g., Kirschner and Tomasello 2010; Reddish, Fischer, and Bulbulia 2013; for a review, see Knoblich, Butterfill, and Sebanz 2011). These affective rewards are associated with different types of synchrony, such as synchronized motor representations (Rizzolatti and Craighero 2004), body postures and gaze patterns (Shockley, Richardson, and Dale 2009), speech patterns (Fowler et al. 2008), facial expressions (Chartrand and Bargh 1999), or heart rate (Vikhoff et al. 2013) of the interactors. A prominent sociological tradition argues that synchronous activities that are ritualistic interactions, such as plays, games, dancing, singing, worship and ceremonies, lack any ulterior purpose or goal and instead have the function of producing affective rewards for the participants (e.g., Collins 2004; Durkheim [1912] 2001; Summers-Effler 2007). The Smith-Sugden hypothesis posits that such rewards are produced via shared emotions and the concomitant fellow-feelings, that is, pleasant awareness of shared emotions. This indeed seems to be the case in rituals, which provide an ample basis for shared emotions that can be experienced at different stages of the ritual: excitement during its performance, joy when it succeeds, fear when its success is threatened, disappointment at failure, anger at those disturbing the ritual or violating its rules, and so on (Helm 2010; Parkinson, Fischer, and Manstead 2005; Salmela 2012).

However, there are many other synchronous behaviors, which are not institutionalized as rituals but nevertheless constitute basic building blocks for our daily interactions with others. These tend to be reflected in experiments where the resulting synchrony is accidental rather than purposefully achieved by the participants: individuals walk, tap fingers, rock chairs, and so on, either in or out of synchrony, or they rate interpersonal rapport between figures or sounds that are presented as in or out of synchrony (e.g., Hove and

¹¹ We are very grateful to an anonymous reviewer for suggesting that the concept of scaffolding might be used to explain their relation.

Risen 2009; Miles, Nind, and Macrae 2009; Wiltermuth and Heath 2009). In these cases, there is little opportunity for feedback and little history of interaction or reason to suppose the presence of shared intentionality. Nevertheless, in the studies cited above, individuals who engaged in synchronous behavior reported greater feelings of interpersonal connectedness, rapport, and affiliation, and they sometimes cooperated more in a social dilemma than people in asynchronous conditions. Because participants also did not report any shared emotions, it is difficult to make sense of such findings on the fellow-feeling accounts (or rather, the fellow-feeling approach is at best silent about the positive effects of such activities). Instead, the social motivation hypothesis seems better equipped in accounting for the pleasure and reward elicited in more basic synchronous activities.

Based on this evidence from different synchrony studies, one might conclude that the fellow-feeling hypothesis and social motivation hypothesis are not in competition but rather explain different *kinds* of synchrony. However, our proposal is that social motivations have a more important role than simply being one alternative explanation. Instead, we contend that they can themselves explain the emergence of fellow-feelings. They do this by providing a basic platform for how we come to anticipate rewards of social interactions, which in turn informs the design and outcome of future interactions. For example, if individuals experience more basic and contingent behavioral synchrony as socially rewarding in its own right, we should also expect them to establish forms of social interaction in which synchrony and rewards are produced *intentionally*. This seems to be precisely what happens in rituals. Although it is very likely that rituals have the capacity to elicit fellow-feelings through the participants' shared emotions at different stages of the ritual, such rewards are nevertheless initially enabled at least in part by social motivations and the anticipation of social rewards.

We thus predict that the opportunity for fellow-feelings, not just in synchronous activities but more generally, is scaffolded by the basic types of social-affiliative rewards described by the social motivation hypothesis. One way of understanding social motivations as a scaffold is in terms of social motivations becoming a condition for more sophisticated interactions like rituals. However, the key point, here, is that in doing so, the social motivations also allow for the emergence of more complex reward structures such as the rewards of experiencing shared emotions. In fact, this also seems to be true of the resentment hypothesis, that is, the desire to meet, and the aversion to frustrate, others' expectations (Sugden). To the extent that these motivations account for prosocial behavior, it is because they are enabled by and elaborated on the more rudimentary social motivations. In fact, we might even have a reason to reconsider moral virtues (Gintis) and the perceived

reasonableness of others' expectations (Bicchieri), as legitimate proximate accounts for (some) prosocial behavior if they analogously seem to depend on the experience and anticipation of social rewards. In short, if our affiliations with other individuals matter, so will their expectations, interests, and moral claims on us. Hence, social motivations explain why moral virtues and others' expectations matter to us or why they, as philosophers like to put it, acquire "motivational force." In fact, it is the strength of the social motivation hypothesis that it can function as a plausible scaffold for a range of different more sophisticated proximate accounts—without any of them necessarily being reduced to the anticipation of an affective reward.

Indeed, given that it is supposed that many social situations in contemporary societies already are governed by some identifiable moral, social, and prudential norms governing behavior (or at least identifiable behavioral interactive patterns), it would be surprising if social motivations were *not* involved in scaffolding the development of more sophisticated preferences such as those described by the existing approaches. In this sense, social motivations are *always there*, so to speak, often by making a prosocial outcome of social interaction—be it actual, anticipated, remembered, virtual—at least, *prima facie* inherently rewarding for us. That is not to say that social motivation could not explain prosocial behavior on its own without some auxiliary motivational account. On the contrary, we have argued with empirical evidence that, in many cases, the motivation to seek social-affiliative rewards is capable of motivating prosocial behavior, for example, in situations where the conditions of other theories of proximate motivation are not met. Thus, the social motivation account stands on its own *in addition to* providing a scaffold for other proximate explanations of prosocial behavior.

6. Conclusion

We began by noting that many existing economic models of prosocial behavior have been rather silent on the *proximate* psychological mechanisms responsible for the target behavior. We identified the most psychologically plausible accounts, and then based on this discussion and convergent recent evidence from several strands of psychological and biological research, we argued for an alternative social motivation hypothesis. Finally, we suggested that the social motivation account provides a plausible scaffold for other more sophisticated proximate motivations that purport to explain prosocial behavior in more complex social interactions.

Let us conclude by briefly addressing some possible implications for economic modeling of prosocial behavior. As we stated at the outset, the empirical and theoretical input from different fields in psychology does not

unequivocally adjudicate between competing economic models of prosocial behavior. Nor have we proposed yet another utility function based on the social motivation hypothesis. Instead, we provided a basic and general proximate account, namely, social motivations that anticipate rewards attached to affiliation and which can scaffold a host of distinct prosocial motivations such as moral virtues, aversion to resentment, pleasure of mutual emotions, and so on.

Although we do not deny that it is possible to compare different models with different motivations in clear-cut experiments (this is one of the points of behavioral experiments, after all), an upshot of our argument is that there probably will not be a general, single utility function that encompasses all prosocial behavior. Instead, our account suggests a form of pluralism that requires different prosocial preferences to be elicited by contextual features such as the nature of the game or interaction, the past history of interaction, the strength of affiliation, and the existing norms or identified teams in a community.

Still, the social motivation hypothesis also suggests some ways to organize this pluralism about preferences. First, given the importance of individual learning history and environment more generally for the precise development of social motivations (as described in both sections 4 and 5), we should expect some considerable heterogeneity among people in terms of their expectations and preferences in experimental and naturally occurring games. A satisfactory theory of prosocial behavior might try to first identify the different *types* of agents in their experiments, and then explain how these different types impact on the interaction such as sustaining or declining levels of cooperation (cf. Gächter 2007). Second, given that social motivations can scaffold a range of distinct prosocial motivations, we should expect that each agent often has conflicting preferences that are elicited by variable salient contextual cues. Thus, a general theory should ideally attempt to identify some general epistemic conditions for the elicitation of different preferences. For example Bicchieri's conditions for payoff transformation and Bacharach's conditions for agency transformation do not necessarily have to be seen as rival hypotheses, but instead they can both be part of a more general hypothesis that predicts that different conditions can elicit different preferences in the same individual. We hope that our proposal has provided some justification for these lines of enquiry.

While what we have said so far is all broadly consistent with the rational choice tradition in psychology and economics (although perhaps suspending some of its ambitions of complete unification), the social motivation hypothesis might also provide a more basic challenge to the framework, which is the following: if affiliative rewards are truly rewards on par with the various

material rewards, it may be unhelpful to define or operationalize rewards solely in terms of monetary payoffs in games, as, for example, social preference models do. After all the research we have discussed suggests that the neurobiological basis for such rewards is both partly distinctive from other reward circuits, including material ones, and arises early in development and is retained throughout life. The hypothesis also makes good evolutionary sense given the importance of social bonding for not only well-being but, indeed, survival. It is unclear if these ends can consistently be operationalized using monetary rewards.

Thus, the social motivation hypothesis really has two primary implications for economic modeling of sociality roles. One is to posit social motivations as a scaffolder that generates motivations for norm-following and team-reasoning, thus offering a candidate proximate or at least developmental explanation of prosocial behavior in the standard sense of representing a choice to forgo immediate material interests. The second and more provocative suggestion is that in many social contexts, including perhaps some experimental ones, the agent is also concerned with anticipated social-affiliative rewards quite independently of the material payoff distribution. In future modeling, it might therefore be worthwhile to try to conceptualize and operationalize social rewards instead of just monetary rewards, for example, by using the kind of basic affiliative stimuli used in the psychological research we have drawn on.

Acknowledgments

Different versions of this article were presented at the following interdisciplinary conferences: British Society for the Philosophy of Science (Exeter), Joint Action Meeting (Berlin), European Network for Social Ontology (Helsinki), European Network for Philosophy of Social Sciences (Venice), and also at Philosophy of Science seminar in Helsinki. We thank all the audiences for their valuable comments and suggestions. We would also like to thank three anonymous reviewers for this journal whose comments have led to a much improved article. Marion has also learnt a lot from co-participants in the interdisciplinary project on social interaction (<http://social-interaction.eu/>): Vivian Bohl, Mog Stapleton, Christoph Teufel, Wouter van den Bos, and Marijn van Wingerden. Michiru wants to thank Chiara Lisciandra for drawing attention to Clavien and Klein (2010) and Lisciandra, Postma-Nilsenová, and Colombo (2013).

Authors' Note

These authors contributed equally to the work and are listed in alphabetical order.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by the Academy of Finland via TINT/Finnish Centre of Excellence for the Philosophy of the Social Sciences. Marion's involvement in the abovementioned project was supported by the Volkswagen Foundation's European Platform for Life Sciences, Mind Sciences, and the Humanities.

References

- Asch, S. E. 1955. "Opinions and Social Pressure." *Scientific American* 193 (5): 31–35.
- Bacharach, M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Edited by N. Gold and R. Sugden. Princeton: Princeton University Press.
- Baumeister, R. F., and M. R. Leary. 1995. "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation." *Psychological Bulletin* 117 (3): 497-529.
- Berridge, K. C. 2003. "Pleasures of the Brain." *Brain and Cognition* 52 (1): 106-28.
- Berridge, K. C., and T. E. Robinson. 2003. "Parsing Reward." *Trends in Neurosciences* 26 (9): 507-13.
- Bicchieri, C. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge, UK: Cambridge University Press.
- Binmore, K. 2006. "Why Do People Cooperate?" *Politics, Philosophy & Economics* 5:81-96.
- Boyd, R., and P. J. Richerson. 2009. "Culture and the Evolution of Human Cooperation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1533): 3281-88.
- Camerer, C. F. 2003. *Behavioral Game Theory*. Princeton: Princeton University Press.
- Chartrand, T. L., and J. A. Bargh. 1999. "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76:893-910.
- Chevallier, C., G. Kohls, V. Troiani, E. S. Brodtkin, and R. T. Schultz. 2012. "The Social Motivation Theory of Autism." *Trends in Cognitive Sciences* 16 (4): 231-39.
- Clavien, C., and R. Klein. 2010. "Eager for Fairness or for Revenge? Psychological Altruism in Economics." *Economics & Philosophy* 26:267-90.
- Coleman, A., B. D. Pulford, and J. Rose. 2008. "Collective Rationality in Interactive Decisions: Evidence for Team Reasoning." *Acta Psychologica* 128:387-97.
- Collins, R. 2004. *Interaction Ritual Chains*. Princeton: Princeton University Press.
- Crews, D. 1998. "The Evolutionary Antecedents to Love." *Psychoneuroendocrinology* 23:751-64.
- Depue, R. A., and J. V. Morrone-Strupinsky. 2005. "A Neurobehavioral Model of Affiliative Bonding: Implications for Conceptualizing a Human Trait of Affiliation." *Behavioral and Brain Sciences* 28 (3): 313-49.
- Durkheim, E. (1912) 2001. *The Elementary Forms of Religious Life*. A New Translation by C. Cosman. Abridged with an Introduction and Notes by M. S. Cladis. Oxford, UK: Oxford University Press.

- Fehr, E., and K. M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 114 (3): 817-68.
- Fowler, C. A., M. J. Richardson, L. M. Kerry, and K. D. Shockley. 2008. "Language Use, Coordination, and the Emergence of Cooperative Action." In *Coordination: Neural, Behavioral and Social Dynamics*, edited by A. Fuchs and V. Jirsa, 261-79. New York: Springer.
- Gächter, S. 2007. "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and Their Policy Implications." In *Economics and Psychology: A Promising New Cross-Disciplinary Field*, edited by B. S. Frey and A. Stutzer, 20-50. Cambridge, MA: MIT Press.
- Gächter, S., and C. Thöni. 2007. "Rationality and Commitment in Voluntary Cooperation: Insights from Experimental Economics." In *Rationality and Commitment*, edited by F. Peter and H. B. Schmid, 175-208. Oxford, UK: Oxford University Press.
- Gamer, M., A. K. Schmitz, M. Tittgemeyer, and L. Schilbach. 2013. "The Human Amygdala Drives Reflexive Orienting towards Facial Features." *Current Biology* 23 (20): R917-18.
- Gilbert, M. 2013. *Joint Commitment: How We Make the Social World*. New York: Oxford University Press.
- Gintis, H. 2010. "Social Norms as Choreography, Politics." *Politics, Philosophy, & Economics* 9 (3): 251-63.
- Gintis, H. 2011. "Gene-Culture Coevolution and the Nature of Human Sociality." *Philosophical Transactions of the Royal Society B: Biological Sciences* 366:878-88.
- Godman, M. 2013. "Why We Do Things Together: The Social Motivation for Joint Action." *Philosophical Psychology* 26 (4): 588-603.
- Guala, F. 2012. "Reciprocity: Weak or Strong? What Punishment Experiments Do (and Do Not) Demonstrate." *Behavioral and Brain Sciences* 35:1-59.
- Guala, F., L. Mittone, and M. Ploner. 2013. "Group Membership, Team Preferences, and Expectations." *Journal of Economic Behavior & Organization* 86:183-90.
- Helm, B. 2010. *Love, Friendship and the Self: Intimacy, Identification, and the Social Nature of Persons*. Oxford, UK: Oxford University Press.
- Heyes, C. 2012. "What's Social about Social Learning?" *Journal of Comparative Psychology* 126 (2): 193-202.
- Hove, M. J., and J. L. Risen. 2009. "It's All in the Timing: Interpersonal Synchrony Increases Affiliation." *Social Cognition* 27 (6): 949-60.
- Kirschner, S., and M. Tomasello. 2010. "Joint Music Making Promotes Prosocial Behavior in 4-Year-Old Children." *Evolution & Human Behavior* 31:354-64.
- Knoblich, G., S. Butterfill, and N. Sebanz. 2011. "Psychological Research on Joint Action: Theory and Data." In *The Psychology of Learning and Motivation*, Vol. 54, edited by Brian Ross, 59-102. New York: Academic Press.
- Lisciandra, C., M. Postma-Nilsenová, and M. Colombo. 2013. "Conformality. A Study on Group Conditioning of Normative Judgment." *Review of Philosophy and Psychology* 4 (4): 751-64.

- Maestro, S., F. Muratori, M. C. Cavallaro, F. Pei, D. Stern, B. Golse, and F. Palacio-Espasa. 2002. "Attentional Skills during the First 6 Months of Age in Autism Spectrum Disorder." *Journal of the American Academy of Child & Adolescent Psychiatry* 41 (10): 1239-45.
- Mason, W., and S. Mendoza. 1998. "Generic Aspects of Primate Attachments: Parents, Offspring and Mates." *Psychoneuroendocrinology* 23:765-78.
- McGuigan, N., J. Makinson, and A. Whiten. 2011. "From Over-Imitation to Super-Copying: Adults Imitate Causally Irrelevant Aspects of Tool Use with Higher Fidelity than Young Children." *British Journal of Psychology* 102 (1): 1-18.
- Merritt, J. M., J. M. Doris, and G. Harman. 2010. "Character." In *The Moral Psychology Handbook*, edited by J. M. Doris, 355-402. Oxford, UK: Oxford University Press.
- Miles, L. K., L. K. Nind, and N. Macrae. 2009. "The Rhythm of Rapport: Interpersonal Synchrony and Social Perception." *Journal of Experimental Social Psychology* 45 (3): 585-89.
- Nelson, E. E., and J. Panksepp. 1998. "Brain Substrates of Infant-Mother Attachment: Contributions of Opioids, Oxytocin, and Norepinephrine." *Neuroscience & Biobehavioral Reviews* 22 (3): 437-52.
- Nielsen, M., and C. Blank. 2011. "Imitation in Young Children: When Who Gets Copied Is More Important than What Gets Copied." *Developmental Psychology* 47 (4): 1050-53.
- Nielsen, M., and K. Tomaselli. 2010. "Overimitation in Kalahari Bushman Children and the Origins of Human Cultural Cognition." *Psychological Science* 21 (5): 729-36.
- Pacherie, E. 2011. "Framing Joint Action." *Review of Philosophy and Psychology* 2 (2): 173-92.
- Parkinson, B., A. Fischer, and A. S. R. Manstead. 2005. *Emotion in Social Relations: Cultural, Group, and Interpersonal Processes*. New York: Psychology Press.
- Paternotte, C., and J. Grose. 2012. "Social Norms and Game Theory: Harmony or Discord?" *British Journal for the Philosophy of Social Science* 64 (3): 551-87.
- Pierce, K., R. A. Müller, J. Ambrose, G. Allen, and E. Courchesne. 2001. "Face Processing Occurs Outside the Fusiform 'Face Area' in Autism: Evidence from Functional MRI." *Brain* 124 (10): 2059-73.
- Reddish, P., R. Fischer, and J. Bulbulia. 2013. "Let's Dance Together: Synchrony, Shared Intentionality, and Cooperation." *PLoS ONE* 8 (8): e71182. doi:10.1371/journal.pone.0071182.
- Rizzolatti, G., and L. Craighero. 2004. "The Mirror-Neuron System." *Annual Review of Neuroscience* 27:169-92.
- Salmela, M. 2012. "Shared Emotions." *Philosophical Explorations* 15 (1): 33-46.
- Schnall, S., J. Benton, and S. Harvey. 2008. "With a Clean Conscience." *Psychological Science* 19 (12): 1219-22.
- Shockley, K. D., D. C. Richardson, and R. Dale. 2009. "Conversation and Coordinative Structures." *Topics in Cognitive Science* 1 (2): 305-19.

- Silk, J. B., J. C. Beehner, T. J. Bergman, C. Crockford, A. L. Engh, L. R. Moscovice, R. M. Wittig, R. M. Seyfarth, and D. L. Cheney. 2009. "The Benefits of Social Capital: Close Social Bonds among Female Baboons Enhance Offspring Survival." *Proceedings of the Royal Society B: Biological Sciences* 276 (1670): 3099-3104.
- Smerilli, A. 2012. "We-Thinking and Vacillation between Frames: Filling a Gap in Bacharach's Theory." *Theory and Decision* 73 (4): 539-60.
- Smith, A. (1759) 2002. *The Theory of Moral Sentiments*. Cambridge, UK: Cambridge University Press.
- Sugden, R. 1993. "Thinking as a Team: Towards an Explanation of Nonselfish Behavior." *Social Philosophy & Policy* 10 (1): 69-89.
- Sugden, R. 2000. "The Motivating Power of Expectations." In *Rationality, Rules, and Structure*, edited by J. Nida-Rümelin and W. Spohn, 103-29. Dordrecht: Springer.
- Sugden, R. 2002. "Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling." *Economics & Philosophy* 18:63-87.
- Summers-Effler, E. 2007. "Ritual Theory." In *Handbook of the Sociology of Emotions*, edited by J. E. Stets and J. H. Turner, 135-54. New York: Springer.
- Tan, J. H., and D. J. Zizzo. 2008. "Groups, Cooperation and Conflict in Games." *Journal of Socio-economics* 37 (1): 1-17.
- Tuomela, R. 2007. *The Philosophy of Sociality*. New York: Oxford University Press.
- Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 185:1124-31.
- Uzgiris, I. C. 1981. "Two Functions of Imitation during Infancy." *International Journal of Behavioral Development* 4 (1): 1-12.
- Vikhoff, B., H. Malmgren, R. Åström, G. Nyberg, S.-R. Ekström, M. Engwall, J. Snygg, M. Nilsson, and R. Jörnsten. 2013. "Music Structure Determines Heart Rate Variability of Singers." *Frontiers in Psychology* 4: 334 (1-16).
- Weatly, T. and J. Haidt. 2005. "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16: 780-784.
- Wiltermuth, S. S., and C. Heath. 2009. "Synchrony and Cooperation." *Psychological Science* 20 (1): 1-5.
- Wise, R. A. 2004. "Dopamine, Learning and Motivation." *Nature Reviews Neuroscience* 5 (6): 483-94.
- Young, A. W., J. P. Aggleton, D. J. Hellawell, M. Johnson, P. Broks, and J. R. Hanley. 1995. "Face Processing Impairments after Amygdalotomy." *Brain* 118 (1): 15-24.

Author Biographies

Marion Godman is a Research and Teaching Associate at History and Philosophy of Science department, Cambridge University and a researcher at the Academy of Finland Centre of Excellence in the Philosophy of Social Sciences (TINT) at Helsinki University. Her research spans a range of topics in the philosophy of science, with a particular interest in the sciences of biology, psychiatry, and psychology.

Michiru Nagatsu is a researcher at TINT, Helsinki University. He writes on philosophy of behavioral and social sciences, with a particular focus on experimental and behavioral economics.

Mikko Salmela is an Academy Research Fellow at Helsinki Collegium for Advanced Studies. He is specialized in interdisciplinarily oriented philosophical study of emotions, and most recently, collective emotions.