Reasonable Trust

Evan Simpson

Abstract: Establishing trust among individual agents has defined a central issue of practical reasoning since the dawning of liberal individualism. Hobbes was convinced that foolish self-interest always threatens to defeat uncompelled cooperation when one can gain by abandoning a joint effort. Against this philosophical background, scientific studies of human beings display a surprisingly cooperative species. It would seem to follow that biologically inherited characteristics impair our reason. The response proposed here distinguishes rationality and reasonableness as two forms of good reasoning. One is consistent with the model of strategic rationality, the other with a model of emotional relationship. From the Hobbesian perspective trusting agents are not rational if their makeup discourages advantageous defection even when one knows it will not be detected or punished. The point is indecisive because reasonable trust insulates cooperative action from the factors that have appeared to make it chancy or unstable without some enforcing power. A critical theme is that trust does not simply rest upon a biological disposition to conform to norms. That would explain but not justify aversion to defection. In fact, trust can survive reasoned challenges to norm-conforming dispositions, displaying the responsible social animal living along with the rational individual.

Establishing trust among individual agents has defined a central issue of practical reasoning since the dawning of liberal individualism. If 'he that performeth first, has no assurance the other will performe after' (Hobbes 1651: 105), how can rational agents trust one another outside a coercive regime? Hobbes was convinced that foolish self-interest always threatens to defeat uncompelled cooperation when one can gain by abandoning a joint effort. In this case it is an empty truism that 'Reasonable trust will require good grounds for . . . confidence in another's good will' (Baier 1994: 99). Since such confidence may always be disappointed, some guarantee of good behavior seems essential for reasonable people to work together in the absence of any assurance of faithfulness.[1]

Against this philosophical background, scientific studies of human beings display a surprisingly cooperative species. We are physiologically inclined to expect well of other persons even in the absence of any personal history and are influenced by concern for equity as much as by personal advantage (cf. Fehr and

Schmidt 1999: 817–68 and Kosfeld *et al*. 2005: 673–6). No doubt trust and cooperation confer selective advantage upon the species as long as the benefits to its members are not outweighed by the danger of being exploited by others. It is then plausible that hormonal readiness to trust and psychological constraints upon self-interest are part of our genetic makeup. In order to understand these suggestions from neuropsychology and experimental economics very fully, however, we need to know whether biologically inherited characteristics impair our reason. The answer I propose distinguishes rationality and reasonableness as two forms of good reasoning. From the Hobbesian perspective trusting agents are not rational if their makeup discourages advantageous defection even when one knows it will not be detected or punished.[2] I grant the point but argue that it is indecisive. The foundations of cooperation are not limited to the Hobbesian rationality that pervades modern philosophy. They also include reasonable emotional trust that mitigates the separateness of strategically rational contractors.[3]

   In developing this case for two alternative grounds for confidence in others, I will first briefly explicate the difference between rational and emotional trust, the contrast between rationality and reasonableness, some relationships between trust and cooperation and some issues about individual and plural agents. In Section 2 I will proceed to show why trust of the strategic kind is unable to account for the most salient feature of human cooperation, namely the offense taken when one defects from a common enterprise. In Section 3 I will characterize emotional trust more fully, describing commitments that contrast with the motivations of Hobbesian agents who move about the world as mutually disinterested strangers unencumbered by second-personal reasons.[4] The contestability of the social norms that are central to this kind of trust makes it reasonable in the requisite way. All of this reinforces my central thesis: the fundamental characteristic of reasonable cooperators is their capacity for emotional relationships rather than their self-interested ability to coordinate their behavior or make conventional agreements.

1. Two Kinds of Trust

'[T]he capacity for social cooperation' in John Searle's view 'is the capacity for collective intentionality' (Searle 2003: 198). If intentions to act are understood as expressions of beliefs and desires, this cannot be quite right. Cooperation is displayed by animals in hives and flocks whose closely coordinated actions consist in cue-driven behavior. The physiological causes of these actions display no separation between belief and desire or perception and response. Rather, perceptions have the biological function of directing specific behavior (Sterelny 2003: 29). By contrast, intentionality requires perceptions to be decoupled, at least in part, from particular behaviors, so that one's beliefs are relevant to a choice among possible actions, depending upon the desires of the agent. In this way, intentional behavior is under the control of the agent's reasons rather than the

environment (cf. Prinz 2004: 45). Collective intentionality is then the capacity for reasoned social cooperation. Understanding this capacity will include understanding the reasons that make it possible for there to be intentional systems consisting of more than one agent. This study focuses upon those reasons, asking when reasoned cooperation is rational and when it is reasonable.

'Rationality' and 'reasonableness' are partially technical notions meant to display some of the variety of reasoned discourse. In particular they will help to explicate the place of reasons in our emotional lives, including social emotions like trust. Rationality has well developed models. It is displayed in both theoretical and practical deliberations, expressing criteria of good epistemic and strategic judgment. Fundamental to these criteria are the truths that support one's beliefs about matters of fact and means to ends, permitting clearly right answers to questions and agreements based in objective evidence. Whereas assignments of rationality are tied to truths in this way, assignments of reasonableness are more permissive.[5] In matters of reasonable belief there need be no single right or best answer to a question. It is because there may be a plurality of reasonable views at the end of the day that the concept is appropriate for emotional response. Emotions have a factual basis but the facts alone do not determine whether an object is worthy of a particular emotion. As a result the evaluative judgments of emotional beings may diverge in ways uncharacteristic of rational observers.

Details of the distinction between rationality and reasonableness will come into sharper focus as we progress, but the framework already permits stating the case that trust should be understood in terms of two families of cognitive capacities, one matching models of epistemic and strategic rationality, the other displaying emotional relationship. Strategic trust is an assessment of another's good will or the likelihood of cooperation. So conceived, it is the subjective probability with which one agent assesses that another will perform a particular action.[6] As an estimation of reliability based on empirical generalizations it may lead to joint action when strategic reasons warrant it, but cooperation of this kind is easily overridden by rational agents whose personal interests come first. In contrast to prudence of this kind, emotional trust provides the connection evident in adverse responses to cheating and betrayal. On the strategic model annoyance is unwarranted because anyone can recognize that when others act contrary to agreements or mutual undertakings their actions represent decisions one could have made oneself in their circumstances.[7] Hence, if contempt of faithlessness and indignation at failures of duty are reasonable responses to disappointed trust, we require a more complex picture of reasoning agents than Hobbesian models of rationality can provide.

The benefits of cooperation for our ancestors explain the emergence of emotional trust, but they do not explicate the reasonableness of the emotion. That requires understanding why adverse responses to cheating and betrayal can be justified within an alternative set of other-regarding reasons. As social animals we are certainly capable of defection and betrayal, but our biology includes a presumption of fidelity. If it can be a reasonable presumption then collective

intentions are to some extent insulated from the features that have appeared to make cooperation chancy or unstable without some superior power to enforce it. The critical point is that if faithfulness rested simply on a biological disposition to conform to norms, then that would explain but not justify aversion to defection.[8] Hence, it is important to see how the reasonableness of emotional trust survives norm-conforming dispositions. We will come to this in Section 3, but need first to sharpen the contrast between trust as a property of rational individuals and trust as a social emotion. A recent collection of original articles edited by Frederick Schmitt helps in this task (Schmitt 2003a).[9] The contributions display the first half of our contrast and the hold of rationality on contemporary philosophical analysis. As Bennett Helm has noted, their authors tend to view collective intentions as ones whose resultant actions are ultimately 'merely those of individuals' (Helm 2008: 21).[10]

This viewpoint is exemplified by Schmitt's own contribution, in which he suggests that, however much we think about collective intentionality as a matter of shared intentions, a correct ontology consists entirely of individuals. Social agency is then conceptually but not ontologically autonomous from individual agency, just as mental agency is sometimes held to be conceptually but not ontologically autonomous from physical agency (cf. Schmitt 2003b). Schmitt's argument depends upon the inevitable disparity between the various beliefs of people engaged in a common enterprise, which supports the inference that there are no plural agents with ends of their own but only individuals coordinating their actions. Philip Pettit's contribution develops the more liberal view that collective intentions should not be modeled too closely on the psychological states of individuals: When an organization determines a course of action its intention is certainly not a mental state, but it is a functional analogue of the beliefs and desires of rational individuals. Yet, even so, the underlying reality is still individualistic in Pettit's view. Well-functioning deliberative groups invite dissenting opinion and if conflicts between the purposes of individuals and the decisions of a group occur too often, some individuals will desert it and the collective agent will prove vulnerable to its individual membership. In such cases people assign priority to their natural persons rather than to their institutional connection, so that the organization loses the authority to assert its own priority.

Pettit equates this capacity of natural persons to decide whether to stand by a plural agent with a natural priority of individuals to groups, thereby suggesting the derivative nature of collective intentionality (Pettit 2003: 190). The equation is defective, though. The independence of individuals as a permanent threat to collective intentions only shows natural persons to have a definable priority to any particular group and its intentions, not to all groups and collective intentions generally. In due course the point will support a distinction between institutional agents and agents who are social in a more robust sense that is lost when questions of constitution are divorced from emotional evaluation, as they are by Schmitt and Pettit.[11] Emotional trust will display the point. It becomes easier to appreciate the fuller sense in which human individuals are also and equally social kinds when the reasonableness of emotional trust is placed in contrast to

assessments of probable behavior. Bernd Lahno has usefully compared such trust with faith, noting that religious faith lacks compelling empirical evidence as its basis (Lahno 2001: 171–89). So, from the standpoint of meta-ethical analysis, does trust. Its reasonableness will not rest upon observable facts alone but upon norms of behavior that support aggrieved responses to perceived infidelity. The social but contestable character of these norms shows how the separateness and union of persons are aspects of a single state of human being.

The fact that the cooperation of birds and bees has little or nothing to do with rationality and reasonableness shows that neither kind of trust is required for cooperation or joint action. Reasoned cooperation does depend upon one, the other or both, but gaining a clear view of the matter requires seeing how evaluative issues intertwine with the question whether groups are constituted by individuals or individuals by groups.[12] The intuition that collective intentions are best analyzed as derivatives of the beliefs and desires typical of individual intentions is stubbornly difficult to confirm, as is the alternative intuition that individuals are naturally emergent from fundamentally plural agents, such as collaborative groups. A comprehensive resolution would require a much larger canvas than this discussion offers. It would include the reality that although trust is often a matter of personal collaboration it need not be. We can entrust much of the management of what we have to governments or persons whom we do not know and who do not know us. This trust may be strategic: 'It is, in any event, efficient and economically justified' (Ijselling 1992: 413). It may be emotional, as when we trust whole professions–politicians, journalists or priests–that prove untrustworthy. Within the scope of this article I only develop the basic personal dimension of trust and the point that collective intentions do not provide good grounds for objecting to defection within the framework of Hobbesian rationality and metaphysics.

Rationality supports cooperation but always with the caveat that personal interests may be pursued over social benefit when the two objectives conflict. Because rule-following and norm-conforming individuals can fail to heed this warning 'the social constitution of individual agency' may refer to nothing more interesting than the instinct to subscribe to conventions even when they collide with one's interests. The class of analyses of joint action and shared intention collected by Schmitt reflects this limitation. A more robust account of constitution will arise through recognizing not only the cooperation emergent from biology alone and bargaining between strangers but also the cooperation characteristic of faithful friends and others who are emotionally responsible to one another. These others are important, for as Andrea Westlund points out they permit extending deeply constitutive relationships beyond friendship to business partners or even complete—but not disinterested—strangers who want to find reasons that apply to them as a group and make them answerable to one another in ways that instrumental concerns alone do not (see Westlund 2009: 15). In keeping with my restricted canvass, I will not attempt to characterize the broader normative state of answerability but use emotional trust to illustrate a matter of relationship that cannot be adequately expressed by conventional or contractual obligations. The

basic lesson is that when cooperation is rational it decomposes into individual intentions but when reasonable it entails discursive connection.

2. Limitations of Rational Trust

If we are doing something together and I break off without warning or explanation you might criticize my action. Understanding what makes my desertion objectionable is a hard problem as long as the relationship between cooperating agents is thought of as resting upon a simple convention rather than an emotional connection. Following David Lewis we can think of conventional behavior as regulated by rules that most people prefer to observe as occasion arises, for example, *Everyone pull on the oars at the same time* or *Drive on the left side of the road*. A convention can be a temporary coordination of behavior or an established practice. In either case it is an arrangement of convenience for the participating agents. They depend upon common knowledge of this sort: everyone has reason to believe that their preferences will be served by a clear regularity of behavior together with reason to believe that others also know this of themselves and ascribe similar knowledge to the others. The result is a self-sustaining pattern of beliefs and preferences. Conventions understood in this way rest upon no obligations other than those arising from rational self-interest. Based upon the parties' objectives the arrangements are open to reassessment leading to consequent abandonment of the venture without going wrong. When one person defects from a stag hunt in order to pursue a rabbit, there is no suggestion in Rousseau's characterization or Lewis's analysis of this encounter that blame can be attached to individuals trying to achieve their separate interests (see Rousseau 1967: 215 and Lewis 1969: 7).

In such circumstances deserters may regret their action if it fails. Fear of failure may deter them from deserting in the first place. These non-social emotions make interesting contributions to cooperative behavior and form part of the larger picture of strategic trust, but they say nothing about the hard problem of assigning blame in such cases. The thriving literature on common knowledge exemplified by Schmitt's volume recognizes this silence as a problem: the no-fault view of collective conventions conflicts with the sense of grievance experienced by those who are disadvantaged by defection from a mutual undertaking. Various responses seek to place joint action within a framework of binding norms that distinguish deontically structured social agency from coordinated individual actions that entail no moral expectations of others' behavior. Success would establish that certain cases of defection from collective intentions include faulty practical reasoning, whose errors are distinct from ignoring threats of punishment that address self-interest alone. The suggested errors include violation of powers of permission, inconsistent joint intentions, incoherent group motivations or contravention of authoritative institutions. After considering each of these suggestions we will have to conclude that the rationalistic framework of strategic trust prevents any practical errors from being

convincingly demonstrated, leaving accounts of collective intentionality too individualistic to justify emotional complaint.

Powers of Permission?

Everyday life includes many mutual engagements. The terms of these encounters are often thought to include a right to expect that one's partner will not withdraw from activity without some form of permission. One might see such a requirement of permissibility as analogous to the justification one needs to provide to oneself when neglecting an intention to act for some individual purpose (see Gilbert 2003 39–64). Other things being equal, failing to follow though upon this intention makes one open to self-criticism and regret. A joint undertaking may similarly be thought to bind persons together in such a way that one is subject to justifiable criticism for abandoning the mutual activity without some form of agreement from one's partner.

Suppose one abandons a walking companion without permission. What exactly justifies complaint and explains what one has done wrong in breaking off the mutual engagement? Distinctively moral wrongdoing is not obviously in question. No clear harm has been done to the other that would warrant anger. No obvious violation of natural rights has occurred that would warrant indignation. There is no clear source of obligation ignored, no promise broken. Possibly, as Thomas Scanlon argues, a principle of fidelity has been violated (Scanlon 1998: 304), but the examination of emotional trust below shows that any such principle and its implied obligation can be credibly challenged. Depending upon the circumstances the action might be prudentially misguided or a lapse of common courtesy, but breaches of foresight and custom do not entail any very strong violation of mutual responsibilities.[13] The only evident error in breaking the engagement is a mismatch of personal intentions. One does not violate a collective intention if one undertakes a joint commitment with a private reservation. One simply realizes that one's partner may be inconvenienced if one finds something better to do, allowing one's own interests to prevail. People then engage as separate individuals who are responsible only to themselves and interested only in maximizing their personal benefits. Consequently, there is no ready answer to the question why one should let one's expressed commitments trump one's desires except to avoid not being trusted in future. In that case the normative force in question is only that of self-interest. Even if we both express the intention to walk or hunt together neither of us is invested with permissive powers stronger than conventional etiquette expects. Any 'wrongness' assigned by disappointed trust then depends upon a dubious intuition of the obligations of plural subjects and their joint actions in a universe of individuals.

Joint Intentions?

The normative dimension of joint action might instead be located in the conditions for certain collective intentions. There are occasions on which another

person's intentions appear to enter essentially into one's practical deliberations, constraining what one can rationally intend. Just as I must organize my desires and beliefs consistently in order to succeed in my private endeavors, you and I must prevent conflicts between our relevant intentional states when there is something we want to do together. Thus, if we have a common objective and know that we can reach it only by taking essential tasks in turn, your intending to begin can determine my intention to go second. If we both have reason to drive non-stop together from Halifax to Boston, your decision to start produces my intention to take over when you need rest. I may, of course, object that we should have discussed the matter before you decided, but if we are committed to our objective then the complaint is peripheral and indecisive. Hence, I may be acting contrary to the requirements of practical rationality if I do not allow your intentions to determine mine in at least some cases of joint action.[14]

Does this account of rational cooperation successfully address Hobbes' problem for the one who performs first? The suggestion depends upon its basic intuition that one can intend another person's actions as well as one's own, but a contrary intuition insists that one can intend only one's own actions. Commands superficially challenge this opposing intuition, since they express one person's intention regarding what another is to do: If my intention to do something settles what I will do, then my command that you do something settles what you will do as long as you consider my order authoritative. Recognizing my normative authority, you think you should comply and form the intention to do as I say. Unfortunately for the first intuition your forming the intention to obey my command is not the same as my intending your action, leaving the proposal unsubstantiated. Explicating the normative demands of joint action in terms of interpersonally fixed intentions therefore remains questionable, especially given the availability of alternative analyses of commands as essentially orders backed by no other authority than threats. In this case, my cooperation in a supposedly joint action may again collapse into prudent concern. No good grounds for offense are identified.

Group Motivations?

The idea that failures of cooperation are irrational when collective intentions collapse can be supported by more realistically complex cases than those considered so far. If collective actions are viewed as linked to multiple decisions over time, the community of agents arguably takes priority: 'If there is a conflict with individual motivation and group motivation the latter wins' (Tuomela 2003: 106). When participating in joint agency one may put private interests aside, sincerely acting for the collective rather than oneself alone, reasoning in the 'we-mode' rather than the 'I-mode.' However, such motivations are not derivable from circumstances of collective intention if the parties begin from self-interest. Consider a series of occasions for choice (A, B, C) when each of a pair of participants must determine whether to act cooperatively by staying in a game ($\rightarrow$) or uncooperatively by withdrawing ($\downarrow$). Either agent is free to leave on any

occasion, at which point the game is over. Their choices are guided by the potential benefits indicated by the rewards (1, 2, 3) in the following diagram, where the respective returns to the first and second persons are given in the brackets.[15]

A   →   B   →   C   →   (2,3)
↓   ↓   ↓
(2,0)   (1,2)   (3,1)
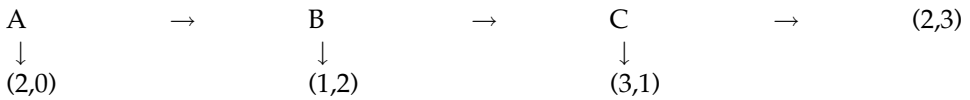
The diagram illustrates the fact that longer periods of cooperation usually yield greater general benefit than interrupted ones: the longer the game continues, the greater the sum of rewards obtained. It also displays circumstances in which the players will be tempted to leave the game in order to improve their personal advantage. Thus, the first player could leave at the first decision point unless there is enough reason to expect the equal or greater payoff available at the third decision point. The second player will want to stay at least until the second decision point (nothing is gained by quitting earlier) but then needs a reason to continue sufficient to overcome concern that the first player will prefer personal benefit over a greater payoff to the group. Such a reason would be the expectation generated by evidence that the first player is prepared to continue as long as the second is. By initially indicating willingness to continue in order to achieve a better result, the first player provides a rationale for mutual cooperation.

Of course, the evidence of the players' willingness to cooperate is very weak here. The evidence becomes strong only over an extended period of engagement, when estimates of the expected benefits warrant continuing rather than ceasing to play. Such estimates include many mature relationships—dealings between merchants and their steady customers, for example. The three-stage model can then serve as a simple representation of this more complicated situation, displaying the fact that there is 'a kind of trust' typical of the players both as modeled and in real life (Cf. Tuomela 2003: 118). Yet this is the kind of trust that rests upon a subjective probability and requires no assumption of fidelity. In coming to the last decision point, estimates of probability will not lead self-interested players to the best collective result, since the first player can be expected to leave the game unless some additional motivation is present. One may be moved by altruistic or utilitarian considerations to value benefits to others or to the group ahead of one's own, but even if these motivations are psychological possibilities nothing in this scenario requires them to be at work. The proposition that group motivation wins in conflicts with individual motivation is therefore not substantiated. Trust as characterized does not lead to completion of the game, since nothing requires the 'we-mode' reasoning that encourages continuing cooperation.

Authoritative institutions?

These attempts to attribute special normative force to cooperative agency are subject to a common difficulty: they are consistent with a view of cooperation as

convention-like. The problem is not fatal if there are conventions that carry legitimate social expectations with them, including rights and responsibilities that can be disregarded only on pain of valid criticism. Searle's constitutive conventions that establish social institutions therefore offer an initially promising remedy.[16] In his view, when dealing with possessions and property, cohabitation and marriage, pieces of paper and dollar bills, we easily distinguish the latter item in each of these pairs according to a rule of the form *X counts as Y*. As individual agents we then inhabit institutional realities whose very possibility depends upon their collective recognition. The institutional facts define rights, duties and authorities that one may accept or spurn, but one rejects them still understanding the force of the constitutive conventions one is breaking. Once possessions historically came to count as property and became governed by rules of acquisition and exchange, dishonest merchants and customers violated well-defined responsibilities, making them open to punishment for abusing the established rights of their trading partners. When workers count themselves as union members, individuals who break solidarity recognize that they are subject to justifiable retaliation from their peers. This is a potentially general account of the deontic power of collective intentions. It might plausibly be extended to defend a need for permission before abandoning an informal undertaking, imply the irrationality of deserting joint intentions and establish a priority of social over selfish agency in games of cooperation.

   If the combined intentions of individuals are counted as institutional facts they may subject persons to criticism for disregarding broad deontic powers, but the criticism cannot go very deep. The assessment of wrongness rests upon constitutive rules that leave institutions without indisputable moral force. In contrast to behavioral conventions, an institution includes rules whose violation is formally improper, but one may still guiltlessly act according to these rules in order to enjoy the benefits of the practice until violating it is more advantageous. One can therefore participate in institutions, as in conventions, while spurning commitment and shrugging off others' anger if discovered. Rejecting the deontic requirements of existing institutions can actually be admirable, as when one condemns slavery or injurious rites of passage as savage practices. There is always a potential discrepancy between conventionally defined deontic powers and the acceptability of the rules constituting them. A constitutive convention can be viewed as a behavioral regularity preferred by enough people to sustain it, but such preferences need not carry any sense of personal obligation to the behavior in question. Resort to institutional facts that entail rights and obligations therefore fails to ground trust and cooperation or justify hostility towards deserters.[17]

   We may conclude that the four above attempts to justify complaints about defection in terms of errors of practical reasoning all fail. It is clear that while rational trust does occur and facilitates cooperation it cannot account for the reasonable admiration we feel for faithfulness to a common task through adversity or the indignation we feel when people violate their social obligations. A successful account requires turning from the framework of subjective

probability that is the home of strategically rational trust to the emotional trust that can make desertion worthy of offense.

3. Emotional Trust

In so far as common-knowledge approaches to collective intentions call upon only the kind of mutual awareness needed for Lewis's conventions, emotional trust remains opaque to reason. These approaches sometimes resemble Kant's idea that the problem of setting up a state can be solved by a nation of devils, or thoroughly calculating individuals whose cooperation can be relied upon when it suits their interests (Kant 1795: 112–13). Contrast Milton's picture of humanity:

> O shame to men! Devil with devil damn'd
> Firm concord holds, men onely disagree
> Of Creatures rational, though under hope
> Of heavenly Grace: and God proclaiming peace,
> Yet live in hatred, enmitie and strife
> Among themselves, and levie cruel warres,
> Wasting the Earth, each other to destroy. (Milton 1667: 496–502)

Devils are unsentimental, incapable of both enmity and cooperative affections towards one another. Each created as a separate species in Christian theology, they could not have evolved instinctive connections with others of their kind. The course of human development led to beings capable of both hatred and trust that includes taking not-so-calculated risks on the basis of amicable relationship. Since feelings of trust do include reasoned judgments—one who is trusted is normally perceived to be trustworthy, for example—they can be reasonable or unreasonable. The task is to show clearly how the reasons in question differ from assessments of probabilities and calculations meant to achieve one's desires.

I will pursue this task in several stages, first by contrasting emotional and factual beliefs in order to characterize the cognitive content of trust and similar emotions. Because this content includes evaluations of emotional appropriateness I next explore the aretaic norms of excellence and deontic norms of obligation that are presented in support of these evaluations. The prominence of aretaic issues in assessing emotional trust reinforces the importance of human relationship but does not show clearly enough what makes this trust reasonable. It is, therefore, important to see that emotional trust is not only social in virtue of the normative standards on which its education depends but also philosophically defensible in virtue of possible reasons for accepting or contesting these standards. Mature emotional trust includes not only the acquisition of norm-conferring standards but also appreciation of their openness to reasonable deliberation.

Emotional Beliefs

Emotional trust is a decidedly social sentiment, comparable in this respect to pity and respect. These feelings display a clear affective and cognitive structure, hence amenability to both physiological arousal and deliberative correction (cf. Jones 1996: 4–25). They evaluate their objects and include related motivational dispositions that may overrule their subjects' other interests. Pity typically attributes undeserved or excessive suffering to a sentient being and includes a wish that the pain be relieved even at inconvenience to oneself. Respect for other persons ascribes worthiness to them that includes the disposition to treat them justly even at one's own expense. In each case the validity of the attribution justifies the emotion and its characteristic motive. A creature that suffers inappropriately is pitiful or worthy of pity and relief. A person possessing inherent value is worthy of respect and requires justice. The desires to assist and to comply with obligations to others clearly have their reasons, but the judgments they express are well characterized only by moving beyond altruism and constitutive conventions to their emotional grounds.

With appropriate caution, these grounds can be explicated in terms of the emotional beliefs that distinguish the cognitive element of social emotions from ordinary factual beliefs.[18] The caution recognizes emotions recalcitrant to beliefs that conflict with them, such as the all-too-common phenomenon of trusting friends or lovers who have a history of disappointing one's trust (cf. D'Arms and Jacobson 2003: 127–45). One may accept a promise, feeling that it will be honored despite thinking that its maker will fail to follow through. If the feeling is said to include a belief, it is one that conflicts with what one knows about the object of trust. While this conflict provides a possible account of the unreasonableness of the emotion, it does so at the cost of identifying a pair of beliefs that seem too transparently contradictory to be psychologically plausible. However, this problem is mitigated by recognizing that conflicts between emotional and factual beliefs are more complex than propositional inconsistency. It can be reasonable to regard another person as untrustworthy even if the other's promises are usually kept, suspecting calculating self-interest as the reason for reliable performance. It can also be reasonable to regard certain people as trustworthy in the main or in a crunch in spite of their tendency to make many casual promises that are lightly disregarded. Beliefs form a family of mental capacities, tracking a variety of commitments that correspond to the difference between rationality and reasonableness. Their conflicts may not be resolvable through factual evidence because emotional beliefs include evaluations and are therefore never equivalent to factual beliefs. Although the latter beliefs serve as reasons for trust, they are not decisive because they do not in general entail trustworthiness, fidelity and other cognate judgments whose content does not represent describable behavior alone.

Richard Holton's alternative to characterizing emotional trust in terms of emotional belief can be used to emphasize how such belief is important for reasonableness (Holton 1994: 63–76). He reminds us that it is one thing to trust

others emotionally and another to rely upon them. Devils can rely on one another without being capable of emotional trust and without feeling resentful if their reliance is disappointed or grateful when others perform as expected. Given this distinction it is plausible to suggest that emotional trust is reliance marked by an appropriate reactive attitude. Analogous analyses are familiar in meta-ethics when emotionally expressive judgments are understood as forms of approval or disapproval evoked by some fact about a condition or event. Thus, most people respond approvingly to the perception of reliability. However, this kind of account neglects a basic feature of emotional judgment. The facts that evoke the reaction are not sufficient by themselves to justify approval or disapproval. They simply cause one to react and do not themselves sufficiently rationalize the response. Perception of a creature's pain may be marked by a form of disapproval, but while the condition may arouse pity it does not make it pitiful. In the absence of a factually adequate translation of that evaluation, an emotional belief about the inappropriateness of the suffering asserts itself as a necessary part of an adequate account. Suffering is pitiful, one wants to say, because it is undue or undeserved. In the case of emotional trust the belief that one is faithful and therefore worthy of trust similarly extends beyond empirical reliability to an evaluative judgment. If the assumption that the judgment is reasonable can be redeemed, trust is not simply reliance plus a reactive attitude.

In defending the reasonableness of emotional trust, it is useful to stress that the content of emotional beliefs is both motivational and evaluative. These beliefs are distinct in both ways from the factual beliefs that fit the strategic model of practical rationality. On the strategic model, intentions result from beliefs and desires that are only conditionally related: if one desires something and believes that certain means are most effective in obtaining it, one has a reason to act. The desires are not fundamentally subject to evaluation by belief, and the beliefs do not by themselves lead to action. Emotional beliefs fit neither element of this model. The first difference is evident in an internal relationship between emotional beliefs and desires: in that emotions, beliefs and desires are not fully decoupled. Rather, in trusting others one's belief in their faithfulness retains its biological function of directing agents towards cooperation. The distinction between factual and emotional beliefs thus maps onto a theoretical difference between motivational externalism and motivational internalism. For the former, beliefs motivate only in conjunction with desires; for the latter, beliefs and perceptions are normally sufficient for action to result.[19] Strategic trust may enable us to achieve the best collective outcome, but it needs some external motivation, such as one player's desire for the other's benefit or for the socially most beneficial outcome. By contrast, emotional trust shares its active principle with its characteristic belief, encouraging cooperative action without reference to desires logically external to it. In this case normative judgment and affective mechanisms are not completely independent systems.[20]

The evaluative content of emotional beliefs marks the other important difference between strategic and emotional thinking and makes instinctive

wishes to cooperate subject to the reasonableness of trusting belief: cooperation is desirable for its own sake only when the one trusted is trustworthy. Judgments of trustworthiness constitute a form of emotional appraisal distinct from the judgment that the probability of someone's performing an action is high enough to take a chance on cooperating with that person.[21] Ordinary factual knowledge lacks the evaluative structure necessary to warrant states that are both cognitive and motivational. Of course, factual knowledge is consistent with the form of evaluation entailed in an estimate of reliability, since an assessment of probabilities can be a straightforwardly empirical characterization. Functional evaluations that reflect something's purposes likewise permit descriptive determination that it is working as it should. Institutional facts also support evaluations: if something is your property I am obliged not to appropriate it without your agreement, although my making that judgment does not entail my acceptance or grasp of moral responsibility. In all of these cases it is possible to make valid evaluative inferences from broadly factual claims. By contrast, the evaluations typical of cognitive-motivational states are not satisfied by such factual modes of confirmation. The judgments about faithfulness or trustworthiness typical of emotional trust require reference to norms that have no sufficient empirical or institutional analysis.

Norms of Evaluation

These norms fall into two different families, corresponding to such broadly moral qualities as fidelity under stress and acceptance of obligations. Fidelity raises aretaic issues having to do with ascriptions of virtue and excellence. It is often an object of pride and admiration, just as faithlessness may be an object of humiliation and contempt. Obligations, by contrast, concern deontic matters of practical necessity. Their familiar emotional signature includes the respect for persons as persons that corresponds to the universal ascription of rights to human beings, together with the indignation and guilt that mark failure to observe these rights (cf. Slote 2001: 4). This is clearly rich territory for theoretical dispute. Virtue theorists often regard deontic concepts as derivative from judgments of goodness and badness, while deontologists may regard virtues as over and above the duties that bind rational agents. My purpose here is not to adjudicate tensions concerning the relative priority of the respective kinds of norms, but I do want to note how the difference between these modes of evaluation has the useful implication of correcting an account of emotional trust suggested by Philip Nickel (2007: 310–12). He distinguishes between predictive trust and moral trust along lines similar to our distinction between strategic and emotional trust, but he suggests that morally trusting a person to do something includes ascribing an obligation to that person to do that thing. Promises are often understood this way. In contrast to intentions they are commitments that cannot be relinquished without penalty.[22] While this is plausible for formal promises, however, breaking a promise is often simply regarded as having

failed to act well—by letting down a friend for example. Disappointment will result, but no penalty is implied except for not being so readily trusted in future.

An important general problem for linking trust and obligation too strongly is that people who reject the implication can nevertheless be trusted as long as they are perceived as faithful in virtue of possessing the character traits of one who can be counted upon. Adhering to this aretaic norm may be admired, but it is not obligatory. It can be construed as an 'operational style' rather than a necessary commitment.[23] It also shows that violations of obligation, like failures to seek permission in Gilbert's account, are not the only way to explain anger and resentment, which can also be aroused by failures of amicable relationship. The limitations of a strictly deontic account are brilliantly expressed by Richard Ford: 'If you trust people unnecessarily, it incurs an obligation on everybody . . . I just decline to make people have to bear extra responsibility for their own insecure intentions.' This decision is supported by the idea that 'We're all separate agents, each underlain by an infinite remoteness' (Ford 2006: 70–1, 343). The aretaic perception can deny this separation, seeing lack of trustworthiness as a weakness of relationship, not as violation of an obligation. When faithfulness lapses the other has a reasonable complaint about a poor friend, companion or fellow agent. It has nothing necessarily to do with deontic constraints.

The limitations of explicating emotional trust through ascriptions of obligation are especially clear in the institutional obligations typical of Searle's constitutive conventions. We have noted that a deontic framework consisting of constitutive rules does define clear rights and obligations, but where commitment to the conventions is lacking there is no perception of practical necessity that does not slide back to prudential rationality or irrational norm-conformity. This difficulty is not avoided by stressing that trustworthiness may occur independently of conventions.[24] Scanlon does this in defending his principle of fidelity, which defines an obligation not to mislead others against their wishes when an assurance of performance has been given in circumstances of mutual knowledge of others' wants and intentions to assure. Approximately, one is obliged to do what one has assured another one would do if that assurance was sought and is being relied upon. It is not possible to state this principle with complete precision because the inevitable contingencies that may override duties of assurance elude exhaustive formulation; but while this is an important fact it is not the serious problem with Scanlon's proposal. The problem is, rather, the liability of the principle to open questions of the sort, 'Are assurances really important?' The conditions of fidelity on Scanlon's account are entirely factual. As a result, their relevance to emotional belief can always be contested, as by Annette Baier's Friends, who trust each other without special assurance always to do what is best and regard expectations of special assurance to be the antithesis of trust (Baier 1994: 170). That they would never offer, 'You assured me', as a demonstration of infidelity illustrates the inevitable gap between descriptive principles and compelling norms of trustworthiness.

Emotional Reasonableness

Whereas strategic trust interests one in being trusted, emotional trust makes one capable of being trustworthy. Clearly, a person who does not value fidelity is unable to credit any rational person with this quality, while those whose cooperation rests upon emotional trust can trust themselves and ascribe trustworthiness to others when it seems reasonable. But when is this? If the tendency to trust is strongly influenced by hormones, then willingness to trust may occur before trust is justified. Our basic question is whether this fact is an embarrassment to reason. It is clearly insufficient to answer that an emotion is defensible as long as no evidence defeats the belief it expresses. In this case emotional beliefs could be accepted by default: trust would be warranted unless the presumption of trustworthiness has been challenged by available evidence. This account does not successfully address Hobbesian concerns about the capacity of others to disguise their ill will and practice deception, opening undefeated trust to too many risks. How, then, can we reasonably trust that we will not be deserted by a traveling companion or disappointed by a public official unless the evidence would justify cooperation in games of advantage played by purely prudential agents?

As part of the perceptual tuning of emotions to the environment there is place for learning and cognitive assessment. Normal adult emotional states seek a defensible equilibrium of feeling and perception. In responding to another person we scrutinize her messages for internal consistency and coherence with other sources of evidence. Such scrutiny is not suspicious surveillance. To the contrary, it is reminiscent of Donald Davidson's principle of charity for understanding others by interpreting as many of their messages as possible as true (Davidson 1984: xvii). At the same time, it requires more than assuming that perceptions of trustworthiness are reasonable as long as they have not been defeated. Some form of actual monitoring is necessary for warranting emotional belief and distinguishing gullibility from epistemic competence.[25] Madame Butterfly's trust of Lt Pinkerton was arguably unreasonable for failing this condition. Of course, we cannot expect trust to demand empirical proof of trustworthiness because no such evidence is sufficient. Identifying alternative grounds will help to display the complex reality within which the relationship of individual agents to one another and their societies can be better discerned.

That one withstands monitoring does not by itself show trustworthiness. It may only show reliability, owing perhaps to inveterate rule-observance that reflects control by the social environment rather than independent agency. How then do we distinguish faithful persons from those who do not deserve our trust? The answer requires a connection between autonomous judgment and social norms, whether these norms are the informal expectations that arise through cooperative interactions or whether they express constitutive rules of the kind that define institutional rights and responsibilities. Learning and adopting these norms is an important part of a child's education. Yet, while much of this learning becomes second nature, reasons for contesting it can be advanced. Although

departing from prevailing norms invites criticism, it is not necessarily deserved because the standards are themselves always subject to challenge, qualification and revision. One example is provided by the sensitivity to novel contingencies that can excuse departing from assurances given. Another is the Friends who reject the standard of assurance. We rightly look for recognition of this contestability in those whom we deem trustworthy, since those who follow rules slavishly have not got their educated instincts under the control of good reasons. It is part of deserving trust that one knows the art of giving acceptable excuses for departing from rigid norms. Another part is being able to recognize that trusting relationships are not bound by tightly specified conventions. To participate in the practice of emotional trust requires recognizing prevailing norms, but mature mastery of the practice means recognizing the contestability of the evaluative concepts that make sound emotional belief possible (whether or not they are actually contested).

The challenges to norms described here are not skeptical doubts about prevailing expectations. Rather, criticism usually aims at clarification and reform, seeking changes to social expectations rather than rejecting them altogether. For example, given conventions of hospitality we can normally be trusted to guard guests in our homes from overt danger but may be uncertain what this means in practice. Should a host be trusted not to let inebriated partiers drive home? An answer can be developed through collective deliberations that redefine customary opinion or the laws of liability governing such matters. To be sure, grounds for trust may remain uncertain because that answer may still be contested by people who have ethical concerns about being one's brother's keeper or political objections to intrusive legal regulation. Reasonable judgments that trust is warranted in particular cases thus express standards of evidence that are never finally established, so that trusting relationships include taking a chance on the adequacy of one's knowledge of others' values as well as their good will. Given people's capacity for amicable self-expression that is part of opening themselves to monitoring, this is not foolish hopefulness.

Emotional trust is, then, distinct from rational trust in resting upon evaluative rather than solely empirical beliefs and judgments. It is nevertheless reasonable when the social norms on which it depends are subject to discursive examination. This leaves only a little more to be said about how reasonable trust promotes the social constitution of individual agency.

We have seen how perceptions of fidelity arise from a first level of social agreement, consisting of innate readiness to cooperate. These trusting instincts are overlain by educated trust, including knowledge of social norms. On the one hand, we human beings naturally form social relationships; on the other hand we may question our norms at a second level, establishing a form of self-assertion. Yet even when reformers and rebels reject conventional norms of behavior and schedules of approval they resemble those who accept external authority naively in assessing the validity of their actions and objectives against a measure independent of their wishes. That they choose or assent to this measure expresses their autonomy, but it shows that such autonomous individuals work within a

conceptual framework that is deeply heteronomous. No one's principles are fundamentally idiosyncratic. We are beings who care about how we are seen by others even when we differ with their standards of evaluation.

Individual autonomy and heteronomy remain distinguished to this extent: the desire for others' approval is not pernicious dependence as long as it expresses the need to be defined by norms one subscribes to thoughtfully. The autonomous individual is then constituted by the ideal community one constructs in this way. This form of constitution lies between the extremes of individuals as norm-conforming creatures whose identities are deeply submerged in their societies and as independent intentional agents who create their social units contractually. It thereby contrasts with Seumas Miller's too strongly individualistic view that 'the social dimension of human actions consists in the regulation of prior individual and interpersonal actions' (Miller 2003: 276). Recognizing reasonable emotional agents as social kinds as well as intentional systems substantiates the anti-Hobbesian point that there is no general way of identifying human actions prior to human relationships.[26] When reasonable trust occurs, in short, we recognize ourselves as beings who are emotionally joined to others rather than as fully independent agents whose social relationships rest insecurely on strategic rationality and coordinated intentions.

As William James noted, one who 'believed no one's word without proof, would cut himself off by such churlishness from all the social rewards that a more trusting spirit would earn' (James 1910: 28). This is not surprising when we realize that most social rewards are not fungible utilities but benefits of reasonable emotional relationships. To be sure, humanity includes disconnected individuals of the kind James described. Not everyone is naturally equipped for emotional trust or other social emotions. Having no intrinsic concern for others' welfare and experiencing no guilt for breaking cooperation, some differ from the vast majority by biological design. Studies of psychopathic personality suggest that when most people are strongly inclined to cooperate, an advantageous niche is available for smaller numbers of defectors (Lalumière *et al.* 2001: 75–92, especially 79). Since the niche is not viable unless its occupants are few, these devils exist only because most individuals display a social nature. Of these exceptional individuals Hobbesian individualism seems right. More-over, for the rest of us brutish circumstances can weaken confidence in others' trustworthiness, interfering with the natural disposition to cooperation. Hobbes was therefore also right to infer that strong social institutions—both informal and legal—are desirable to discourage this tendency, except that the enforce-ment of institutional obligations is not well secured through fearful power. Cooperation is better encouraged through the cultivation of trust rather than fear because the social emotions connect reasonable human beings. If the connections are fragile they are also deep. The coupling of emotional per-ceptions and motivations explains how our biology can govern our actions without compromising our reason. Distinguishing clearly between emotional reasonableness and strategic rationality thus displays the social animal living along with the rational agent.

Conclusion

The Hobbesian picture of trust and cooperation is seriously incomplete. Strategic thinking can lead to enduring agreements but the permanent possibility of deserting those arrangements poses a challenge that is largely met through emotional relationships. By supplementing the standard account of rational individuals with reasonable agents predisposed to trust others we weaken the modern tendency to think of human beings as inherently separate persons whose deliberate negotiations are sufficient to end their isolation and keep them together under a common power. Social relationships precede individual autonomy, but this does not mean subjection to political or customary authority. The spell of social authoritarianism can be broken by appreciation of the capacity for emotional reasonableness that distinguishes most human beings from other collective agents and enables our enjoyment of cooperation without being constantly tempted by the further benefits of breaking faith.

Evan Simpson
Department of Philosophy
Memorial University of Newfoundland
Canada
esimpson@mun.ca

NOTES

[1] As Carolyn McLeod points out, in order to understand trust we must understand good will more as Kant does than as in the colloquial sense of kindly feelings. See McLeod 2002: 21–2.

[2] 'Hobbesian' rather than 'Hobbes'. Philip Pettit plausibly questions whether this answer is faithful to Hobbes's actual texts, which stress the foolishness of departing from laws of reason. Cf. Pettit 2008: 112. The present paper explores the conceptual issue whether advantageous defection is acceptable if one knows it will not be detected or punished.

[3] Alternatively, one might speak of 'a distinctive kind of rationality governing the interconnections among our felt evaluations and evaluative judgments . . .' as Bennett Helm does in Helm 2001: 199.

[4] For this nice characterization see LeBar 2009: 655.

[5] They depart from 'Rational Uniqueness', the view that 'there is a unique maximally epistemically rational response to any given evidential situation'. For this characterization see Christensen 2007: 210.

[6] For this standard definition of 'trust' see Gambetta 1988: 217.

[7] In discussing the trustworthiness of testimony, Paul Faulkner makes a similar distinction between predictive and affective trust. See Faulkner 2007: 312.

[8] For such a norm-conforming view see Heath 2008.

[9] One notable omission from this collection is work by Michael Bratman. The present paper limits itself to the writers on mutual awareness collected by Schmitt. Of these, it omits discussion of several articles on the philosophy of language that lie beyond its scope.

[10] Helm includes Bratman under this judgment.

[11] Another recent attempt to deal with questions of metaphysical constitution in terms of joint intentions considered independently of emotional connection is made by Alonso 2009: 444–75.

[12] Charles S. Peirce provides a classic statement of this issue in Peirce 1992: 105.

[13] These problems are confirmed in a review Gilbert 2006. See Miller 2008.

[14] For such an argument see Roth 2003: 65–9.

[15] The diagram is adapted from Tuomela 2003: 117.

[16] Searle maps this territory in Searle 2003 as well as in other recent writings.

[17] 'Constitution', it is right to agree with John Haugeland, 'can be essentially more than mere institution'. See his Haugeland 1998: 318.

[18] Nothing much hangs upon the expression 'emotional belief'. Alternatively one might refer to emotional 'appraisal' as recommended by Prinz 2004: 50–1.

[19] The qualification, 'normally', recognizes that for purposes of analysis this characterization can be stated more precisely. Cf. Simpson 1999: 208–9.

[20] Contrast with Nichols 2004: 29.

[21] Contrast with Gambetta 1998: 218.

[22] Thus Brandom 1994: 264.

[23] Cf. Calhoun 2009: 613–41, where she argues that shaping one's life around one's commitments is an optional style for managing one's existence (615).

[24] Cf. Westlund on the 'dialogical sensitivity' of 'co-deliberators whose answerability to one another 'is not to be understood as an unconditional obligation to follow through', Westlund 2009: 10, 16.

[25] For a similar point see Henderson 2008: 190–221.

[26] Much of the extensive literature on this action-theoretic position is reviewed by Sneddon 2006.

REFERENCES

Alonso, F. M. (2009), 'Shared Intention, Reliance, and Interpersonal Obligations', *Ethics*, 119: 444–75.

Baier, A. (1994), 'Trust and Antitrust', in A. Baier, *Moral Prejudices*. Cambridge, MA: Harvard University Press.

Brandom, R. (1994), *Making It Explicit: Reasoning, Representing and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Calhoun, C. (2009), 'What Good Is Commitment?', *Ethics*, 119: 613–41.

Christensen, D. (2007), 'Epistemology of Disagreement: The Good News', *Philosophical Review*, 116: 187–217.

D'Arms, J. and Jacobson, D. (2003), 'The Significance of Recalcitrant Emotions', in A. Hatzimoysis (ed.), *Philosophy and the Emotions*. Cambridge: Cambridge University Press.

Davidson, D. (1984), *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Faulkner, P. (2007), 'A Geneology of Trust', *Episteme*, 4: 305–21.

Fehr, E. and Schmidt, K. M. (1999), 'A Theory of Fairness, Competition, and Cooperation', *The Quarterly Journal of Economics*, 114: 817–68.

Ford, R. (2006), *The Lay of the Land*. New York: Alfred A. Knopf.

Gambetta, D. (1988), 'Can We Trust Trust?', in D. Gambetta (ed.), *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell.

Gilbert, M. (2003), 'The Structure of the Social Atom: Joint Commitment as the Foundation of Human Social Behavior', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

—— (2006), *A Theory of Political Obligation*. Oxford: Oxford University Press.

Haugeland, J. (1998), *Having Thought*. Cambridge, MA: Harvard University Press.

Heath, J. (2008), *Following the Rules: Practical Reasoning and Deontic Constraint*. Oxford: Oxford University Press.

Helm, B. (2001), *Emotional Reason*. Cambridge: Cambridge University Press.

—— (2008), 'Plural Agents', *Noûs*, 42: 17–49.

Henderson, D. (2008), 'Testimonial Beliefs and Epistemic Competence', *Noûs*, 42: 190–221.

Hobbes, T. [1651] (1965), *Leviathan*. Oxford: Oxford University Press.

Holton, R. (1994), 'Deciding to Trust, Coming to Believe', *Australasian Journal of Philosophy*, 72: 63–76.

Ijselling, S. (1992), 'Time and Space in Technological Society', *Man and World*, 25: 409–19.

James, W. (1910), *The Will to Believe and Other Essays in Popular Philosophy*. New York: Longmans, Green, and Co.

Jones, K. (1996), 'Trust as an Affective Attitude', *Ethics*, 107: 4–25.

Kant, I. [1795] (1991), 'Perpetual Peace: A Philosophical Sketch', in I. Kant, *Kant: Political Writings*, 2nd edn, trans. and ed. H. S. Reiss. Cambridge: Cambridge University Press.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. and Fehr, E. (2005), 'Oxytocin Increases Trust in Humans', *Nature*, 435: 673–6.

Lahno, B. (2001), 'On the Emotional Character of Trust', *Ethical Theory and Moral Practice*, 4: 171–89.

Lalumière, M. L., Harris, G. T. and Rice, M. E. (2001), 'Psychopathy and Developmental Instability', *Evolution and Human Behavior*, 22: 75–92.

LeBar, M. (2009), 'Virtue Ethics and Deontic Constraints', *Ethics*, 119: 642–71.

Lewis, D. (1969), *Convention: A Philosophical Study*. Oxford: Blackwell.

McLeod, C. (2002), *Self-Trust and Reproductive Autonomy*. Cambridge, MA: MIT Press.

Miller, D. (2008), 'Review of Margaret Gilbert, *A Theory of Political Obligation*', *The Philosophical Quarterly*, 58: 755–7.

Miller, S. (2003), 'Individual Autonomy and Sociality', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield.

Milton, J. [1667] (2007), '*Paradise Lost*, 2', in J. T. Shawcross and M. Lieb (eds), *An Authoritative Text of the 1667 Edition*. Pittsburgh: Duquesne University Press.

Nichols, S. (2004), *Sentimental Rules*. Oxford: Oxford University Press.

Nickel, P. J. (2007), 'Trust and Obligation-Ascription', *Ethical Theory and Moral Practice*, 10: 309–19.

Peirce, C. S. (1992), *The Essential Peirce: Selected Philosophical Writings Vol. 1*, eds. N. Houser and C. Kloesel. Bloomington and Indianapolis: Indiana University Press.

Pettit, P. (2003), 'Groups with Minds of Their Own', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

—— (2008), *Made with Words: Hobbes on Language, Mind, and Politics*. Princeton: Princeton University Press.

Prinz, J. J. (2004), *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Roth, A. S. (2003), 'Practical Intersubjectivity', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

Rousseau, J.-J. (1967), *The Social Contract and Discourse on the Origin of Inequality among Mankind*, ed. L. G. Crocker. New York: Washington Square Press.

Scanlon, T. M. (1998), *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schmitt, F. (ed.), (2003a), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

—— (2003b), 'Joint Action: From Individualism to Supraindividualism', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

Searle, J. R. (2003), 'Social Ontology and Political Power', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

Simpson, E. (1999), 'Between Internalism and Externalism in Ethics', *The Philosophical Quarterly*, 49: 201–14.

Slote, M. (2001), *Morals from Motives*. Oxford: Oxford University Press.

Sneddon, A. (2006), *Action and Responsibility*. Dordrecht: Springer.

Sterelny, K. (2003), *Thought in a Hostile World: The Evolution of Human Cognition*. Oxford: Blackwell.

Tuomela, R. (2003), 'The We-Mode and the I-Mode', in F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield Publishers.

Westlund, A. C. (2009), 'Deciding Together', *Philosophical Imprint*, 9: 1–17.