Economics and Philosophy, **21** (2005) 15–32 Copyright © Cambridge University Press
doi:10.1017/S0266267104000367

CONSTRUING SEN ON COMMITMENT

Philip Pettit
Princeton University

One of Amartya Sen's most distinctive claims, and perhaps also one of his most controversial, is that there is an altruistic attitude toward others that does not make sense within the terms of rational choice theory, however broadly that theory is construed. He describes this attitude as one of "commitment" to others, contrasting it with egoism and also with a distinct altruistic attitude he calls "sympathy." It involves recognizing the goals of another and, regardless of whether or not this answers to independent goals of one's own – regardless of whether or not one internalizes those goals in "sympathy" with the other – letting them impact on how one behaves.

Committing oneself to another in this way is, on the face of it, a fairly common sort of exercise. Surely we often do take the goals of others into account in the manner proposed. We are bent on pursuit of our own ends, we discover that acting as they require will frustrate someone else's projects, and then we pause, take stock, and adjust what we were going to do so that the other will not suffer unduly at our hands. But though commitment is phenomenologically plausible in this way, it looks to be architecturally problematic. On Sen's conceptualization, it characteristically involves putting aside one's own goals and acting on those of another. But how could one ever fail to act on one's own goals? The idea offends against a picture of human psychology accepted on almost all sides, whether among professionals or among the folk.

This paper attempts to expound Sen's view, especially as he has presented it in the most recent statements that he has given or reprinted (Sen 2002); and then to explain why he is led to endorse a position that offends, in formulation if not in substance, against our common sense about action. There are two parts to the paper. The first is devoted to the expository task, the second to the explanatory one.

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

1. EXPOUNDING SEN

1.1 Varieties of rational choice theory

The problem that leads Sen to make a distinction between sympathy and commitment arises within the context of rational choice theory, as he himself calls it. On a minimal understanding, rational choice theory amounts to nothing more than the claim that human behavior – as it materializes, presumably, in the absence of "perturbing" factors and within "feasible" limits – is regular enough to be representable as "maximizing behavior with an identifiable maximand" (Sen 2002: 30). Maximizing behavior will never lead to choices that are seen as worse than available alternatives, though it does not require the choice of the "best" option, as in "optimization"; no "best" option may be definable due to incompleteness in the agent's preference-ordering (Sen 2002: 181).[1]

Sen contrasts with this minimal understanding of rational choice theory the interpretation under which the maximand is "interpretable as the self-interest of the person." Such self-interest, he says, may in turn be understood in either of two ways: in a "narrowly self-centered" sense or in the sense in which my self-interest may encompass the welfare of others, being sensitive to how certain others fare, not just to how well I do myself (Sen 2002: 30–31).

The distinction between the two ways of understanding rational choice theory corresponds to a distinction between two ways of understanding utility in the standard theory of preference and choice. In the one approach, corresponding to the minimal version of rational choice theory, "'utility' is simply defined as the maximand that a person can be seen as promoting"; in the other, corresponding to the self-interest versions of the theory, "'utility' is used as also representing the person's self-interest or well-being" (Sen 2002: 27).

The distinction between different ways of understanding rational choice theory also maps onto an account that Sen gives of how the self may be implicated in choice. To believe in the minimal version of the theory is simply to believe that one's choices are based on one's own goals; it is to believe in what he calls "self-goal choice." To believe in the other version is to hold, much more demandingly, that one's choices are based on the goal of maximizing one's own welfare – one's self-interest – whether that be understood narrowly or broadly: whether it be understood just

[1] Following Sen (2002: chs 3–4), I shall understand it in a broad sense that allows what he calls "menu-dependence" and related phenomena. Menu-dependence means that an action, A – say, taking an apple from a bowl of fruit offered by another – varies in its identity as an object of preference and choice, depending on context; it will be a polite action if there is an apple left, impolite if there is not (Pettit 1991).

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

as welfare, no matter how affected, or as self-centered welfare (Sen 2002: 33–34, 213–14).

1.2 Sympathy and commitment

With these distinctions in place, Sen characterizes two forms of altruism, sympathy and commitment. Sympathy materializes when one pursues one's own self-interest but that self-interest is positively sensitive to the welfare of others, say because one is a generous, affectively responsive sort of agent. It stands opposed to antipathy, which occurs when one's self-interest is negatively sensitive to the welfare of others, say because one is a begrudging, envious type of person, or because one is feeling resentful. "Sympathy (including antipathy when it is negative) refers to 'one person's welfare being affected by the state of others' (e.g., feeling depressed at the sight of the misery of others)" (Sen 2002: 35; cf. 214). Sympathy comes about via enlarging self-interest to the point where it encompasses others too. Thus Sen (2002: 177) says: "Altruism through sympathy is ultimately self-interested benevolence." It is not self-interested in the sense of being pursued with an instrumental eye to securing some personal benefit. It is self-interested in the sense that the person we favor is someone whose welfare matters to us, intuitively, in the same manner as our own; let them fare well and we feel good, let them fare badly and we feel bad.

Altruism through commitment is meant to contrast with altruism through sympathy. The core idea in commitment, endorsed by Sen (2002: 214) over many years, is that it breaks "the tight link between individual welfare (with or without sympathy) and the choice of action." For example, it may involve "acting to help remove some misery even though one personally does not suffer from it." Sen (2002: 177–78) traces the idea of commitment to Ragnar Frisch but finds it already present in Adam Smith. "Doing things for others through commitment may require one to 'sacrifice some great and important interest of our own', as Adam Smith puts it in distinguishing 'generosity' from 'sympathy'."

Where sympathy enlarges a person's self-interest, commitment transcends it. Where sympathy transforms the motor of self-interest, tuning it to the welfare of others, commitment puts another motor in its place. In sympathy, so the idea goes, one's sentiments resonate in common with others so that in acting on the basis of those sentiments one naturally takes account of the welfare of others. In commitment one does not resonate in that manner, or at least one may not do so. Rather, what happens is that while one's sentiments push one in this or that direction, the recognition that others will suffer as a result of going in that direction causes one to alter trajectory. Without relying on the warm stirrings of fellow-feeling, the cold, clear light of reason leads one to change tack.

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

1.3 Two kinds of commitment

How exactly does this happen, according to Sen? He identifies two possibilities and I will describe these respectively as goal-modifying and goal-displacing commitment. He is quite clear about the distinction but does not give the two forms explicit names (Sen 2002: 35, 214).

Goal-modifying commitment occurs when I recognize the goals of others, see that they will be negatively affected by what I am about to do, and alter my own goals as a result; in particular, alter those goals without undergoing the transformation of sentiment that sympathy would involve. My modified goals will reflect, not self-interest, however enlarged, but rather "broader values" that bear on how others are to be treated or how common goals are to be promoted. There is no naturalistic mystery involved in thinking that values of this kind may matter to us, according to Sen (2002: 25), since selectional pressures are liable to have induced in us a concern for such things. And it is plausible in any case, he says, that the capacity for "reasoning and self-scrutiny" – the capacity with which Sen (2002: 4) ultimately identifies rationality – can transform our goals in this way. "Our choices need not relentlessly follow our experiences of consumption or welfare, or simply translate perceived goals into action. We can ask what we want to do and how, and in that context also examine what we should want and how. We might or might not be moved by moral concerns or by social reasons, but neither are we prohibited from entertaining these questions, in shaping our values and if necessary revising our objectives in that light" (Sen 2002: 36).

The acknowledgement of goal-modifying commitment does not require a serious departure from rational choice theory, at least on the minimal interpretation of that theory. Even as I become committed in this way, I will conform to rational choice theory in the minimal sense. I will continue to promote my modified goals – I will maximize in the familiar pattern (Sen 2002: 37) – though the goals I come to serve will no longer be the goals of self-interest, even enlarged self-interest. I may not maximize explicitly, since some of the values by which I am moved will take the form of constraints on how to behave (Sen 2002: 214). But acting on such constraints will still count as maximizing so far as the constraints can be "incorporated into a suitably broadened maximand" (Sen 2002: 41).[2]

Goal-modifying commitment is less radical in this respect, however, than the second, goal-displacing variety of commitment. In this variety, which bulks larger in Sen's discussions, the recognition of the goals of others does not lead me to modify my own goals but rather to displace them. It leads me to take my guidance, not just from my own aims, but also from the goals that I see those others espouse. Commitment, as Sen

[2] The maximand will take account of menu-dependent preferences (Sen 2002: 178).

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

(2002: 35) says, "can alter the person's reasoned choice through a recognition of other people's goals beyond the extent to which other people's goals get incorporated within one's own goals."

Where goal-modifying commitment does not require a departure from rational choice theory, as minimally understood, goal-displacing commitment certainly does. According to the minimal version of rational choice theory, people can be represented in action as maximizing an identifiable maximand, or as acting on their own goals: satisfying the assumption, as Sen calls it, of "self-goal choice." But Sen (2002: 215) maintains that people may be committed to others in such a way that they no longer act in this way on their own goals; "the pursuit of private goals may well be compromised by the consideration of the goals of others." People may become the executors of a goal-system that outruns the private goals that they endorse in their own name: a goal-system that makes place for the goals of others or for the goals of groupings in which people cooperate with others.

This claim is highly implausible, at least on the face of it. Rational choice theory in the minimal sense is close to common sense. It picks up the assumption that when we act intentionally, then we try to advance certain goals in a way that is sensible in light of the apparent facts (Pettit 2002: Essay II.2). The claim that we can be the executors of a goal-system that outruns our own goals is bound to raise a question. Sen (2002: 214), indeed, acknowledges the fact. "It might appear that if I were to pursue anything other than what I see as my own 'goals', then I am suffering from an illusion; these other things are my goals, contrary to what I believe."

Sen's response to this problem is characteristically imaginative, though I don't think it is very satisfying. He says that apart from acting in a way that reflects my real goals I can also act in a way that reflects imagined or as-if goals; I can act as if the goals of others, or the goals of a group to which I belong, are my own. One context in which this occurs is when I assume the responsibility of a trustee, as it were, seeing myself as charged with furthering the other's interests (Sen 2002: 179). Another is that in which I see myself as the representative of a group, charged with doing as well as possible by its interests. People, Sen (2002: 214) thinks, might use this representative identity to get out of game-theoretic predicaments where acting on their separate preference orderings would lead to a bad result for each: "if people are ready to act (individualistically) on the basis of some 'as if' – more cohesive – orderings, then they can do better than acting individualistically in direct pursuit of their real goals."

In stressing these possibilities, Sen suggests that apart from sympathizing with others and escaping from a self-centered version of self-interest, people can also "identify" with one another – adopt the goals they share with others on an as-if basis – and achieve a much more radical liberation.

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

They can escape, not just the rule of self-interest, however sympathetically enlarged, but even the rule of their own private goals. As Sen (2002: 216) himself puts it, "the sense of identity takes the form of partly disconnecting a person's choice of actions from the pursuit of self-goal." He believes that this disconnection occurs on a more or less routine basis in any society. "One of the ways in which the sense of identity can operate is through making members of a community accept certain rules of conduct as part of obligatory behavior toward others in the community. It is not a matter of asking each time, What do I get out of it? How are my own goals furthered in this way?, but of taking for granted the case for certain patterns of behavior toward others."

1.4 The problem with goal-displacing commitment

The notion of goal-displacing commitment remains problematic, despite Sen's attempts to make it plausible by reference to as-if goals and the idea of identifying with another. First of all, I see no real difficulty in making sense of what it is to act in a representative or trustee role – or in any role of the kind – within the terms of self-goal choice. And second, I find it hard to see how one can seriously envisage giving up on the idea of self-goal choice that goal-displacing commitment is supposed to violate.

I am myself quite well disposed toward the idea that we each often act in the name of goals that are endorsed in common with others, as when we represent a group; I hold indeed that we may even have to act on the basis of judgments that we do not ourselves support (Pettit 2003). But the goals endorsed in common with others are still goals we each endorse, and so are in that intuitive sense *our* individually endorsed goals. I individually have it as a goal that we do so and so, you have that as an individually endorsed goal, and so has each of us in the group. True, I will not be able to realize my individually endorsed goal that we do so and so without the help of others. But that does not mean that it is not my goal. We might as well say that because my success in an archery competition depends on the wind, the goal of hitting the bull's eye is not my individually endorsed goal when I take aim at the target.

What is true in the case of acting for the goals of a group, in the role of representative, is also true of acting for the goals of another as a trustee or advocate or good Samaritan, or acting for a goal that represents a compromise between the other's goals and my pre-existing objectives. The goals adopted do not mainline my mind, as it were, and take possession of it. Rather they become goals that I take over as my own, even if they are goals such that, like hitting the target in the archery competition, I am not entirely in control of their realization; their realization may depend equally on how the other behaves.

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

These lines of thought suggest that acting as a representative or a trustee or whatever can make perfect sense within rational choice theory, minimally interpreted; it can be a sort of action that fully respects self-goal choice. That alone gives reason to be surprised at the position Sen takes. But what really makes his position quite problematic is that it is very hard to see how one can give up on the idea of self-goal choice that rational choice theory incorporates. For this idea is close to the core of the common sense psychology of which minimal rational choice theory is an explication (Pettit 2002: Essay II.2).

According to that shared, folk psychology intentional action is a form of behavior that is controlled by the agent's desire to realize a certain condition – the desired condition will count as the agent's goal – and by the agent's beliefs about how best to do that. The goal represents the success condition of the action and will be discernible in how the agent is disposed to adjust the behavior as circumstances change and as different interventions are clearly needed for the realization of the condition. To imagine an action that is not controlled by a goal of the agent, by the lights of this approach, will be like trying to imagine the grin on the Cheshire cat in the absence of the cat itself. Let the agent not have a goal and it becomes entirely obscure how the agent can be said to act; to act, or at least to act intentionally, is to act with a view to realizing a goal.

When Sen alleges that goal-displacing commitment takes the agent beyond the control of his or her goals, then he is setting himself against our basic common sense about action. What rational choice theory asserts on the minimal interpretation is that rational agents act out of a concern for maximizing the expected realization of their goals. And that is precisely to argue, as in our common-sense psychology, that rational agents aim at satisfying or fulfilling their desires, according to their beliefs. It amounts to nothing more or less than asserting the soundness of the belief-desire schema.

Why might Sen be led to question the soundness of that very basic schema? What limitation does he find in the rational-choice representation of action, even when minimally interpreted, that prompts him to think that it is not a comprehensive picture of human decision-making? This is the question to which the remainder of my paper will be devoted.

2. EXPLAINING SEN

In discussing the requirements of rationality, and in making room for departures from the narrower assumptions of economics, Sen often talks of answering to the demands of reason, subjecting our choices to reasoned scrutiny, being guided by broader values and objectives, and the like (Sen 2002: ch. 1). All of this suggests that in his view the belief-desire schema of folk psychology – equivalently, the goal-maximization schema of rational

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

choice theory – is too restrictive to give us the full truth about human decision-making. It does not do justice to the diversity of "reasons for what one may sensibly choose" (Sen 2002: 6). That is why he is led to recognize the possibility of a goal-displacing commitment in which, as it seems, people altogether transcend the range of possibilities envisaged in folk psychology.

There are three possible explanations, I think, as to why Sen might take the view that folk psychology is over-restrictive in this way. Quite independently of how plausible they are as explanations of his line – I shall be suggesting that the first two are implausible – they are of some interest in themselves: they point to different ways in which folk psychology and rational choice theory may be represented and, I would say, misunderstood. I review them, turn by turn, in this second part of the paper.

The three explanations are linked respectively with three theses and my discussion will focus on these.

1. *The no-deliberation thesis*. The schema of belief and desire – and the rational choice theory that seeks to explicate it (Pettit 2002: Essay II.2) – does not make room for deliberation at all. Thus, when one operates in accord with that schema one cannot be deliberating and when one deliberates – assuming deliberation occurs – one cannot be acting in accord with that schema.

2. *The selfish-deliberation thesis*. The schema does make room for deliberation but the only deliberation accommodated is of an inherently selfish variety. Thus, when one operates in accord with that schema one can be deliberating but only in the selfish manner; and when one deliberates non-selfishly – assuming this sometimes occurs – one cannot be acting in accord with that schema.

3. *The integrated-deliberation thesis*. The schema allows for non-selfish deliberation, but only deliberation from a limited basis: that of goals that the agent has internalized and integrated into a standing structure. Thus, when one operates in accord with the schema one can only be deliberating on that limited, integrated basis; when one deliberates otherwise – if this ever happens – one cannot be acting in accord with the schema.

2.1 The no-deliberation account

According to the belief-desire schema rational agents act so as to promote the satisfaction of their desires according to their beliefs; this is the fundamental idea preserved in rational choice theory, minimally interpreted. But one striking fact about the schema, so understood, is that it does not require anything on the part of agents that we might be inclined to call deliberation.

For all that the picture says, desires are goal-seeking states, beliefs fact-construing states, and what happens in action is that they combine to

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

produce the behavior in question. But they can combine to do this without any extra intentional activity – driven in turn by beliefs and desires – of deliberating over whether the goals are appropriately sought, the facts appropriately construed, or the behavior appropriately selected in light of those beliefs and desires. The goal-driven, fact-sensitive creature may be a more or less autonomic mechanism or organism.

It is for this reason that on most accounts, folk psychology is taken to apply to non-human animals as well as to human; in particular, to non-human animals who show no signs of being able to reason. Under this picture, non-human animals are tuned by evolutionary and experiential pressures so that in appropriate circumstances they will act for the realization of certain goals and, in particular, will act in a manner that makes sense under the way they take the facts to be: under the representations of the environment – the more or less reliable representations – that their perceptions and memories evoke. Such animals will instantiate goal-seeking and fact-construing states and those states will interact in such a way as to produce suitable behavior. The animals will be rational agents in the sense of conforming to the minimal version of rational choice theory. Or that will be so, at any rate, in the absence of intuitively perturbing influences, within intuitively feasible limits: for short, in normal conditions.

But if folk psychology is as likely to be true of various non-human animals as it is of creatures like us, there is still a yawning divide between how we and they manage to conform to this psychology (Pettit and Smith 1990). We do not just possess beliefs and desires in the manner of non-humans, and act as those states require. We can give linguistic expression to the contents of many of those states – we can articulate the goals sought and the facts assumed. We can form higher-order beliefs about those goals and facts; beliefs, for example, to the effect that certain forms of consistency or coherence or mutual support do or do not obtain amongst them. And we can seek out such higher-order beliefs with a view to maximizing the checks on the overall pattern of attitudes that is going to unfold within us (McGeer and Pettit 2002).

The exercise we undertake in seeking out higher-order beliefs with the aim of increasing the checks on our overall attitudes is easily illustrated. Suppose I find myself prompted by perception to take it to be the case that *p*, where I already take it to be the case that *r*. While my psychology may serve me well in this process, it may also fail; it may lead me to believe that *p*, where "*p*" is inconsistent with "*r*". But imagine that in the course of forming the perceptual belief I simultaneously ask myself what I should believe at a higher-order level about the candidate fact that *p* and about the other candidate facts I already believe. If I do that then I will put myself in a position, assuming my psychology is working well, to notice that the alleged fact that *p* and the alleged fact that *r* are inconsistent, and

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

so my belief-forming process will be forced to satisfy the extra check of being squared with this higher-order belief – a crucial one, as it turns out – before settling down.[3]

In this example, I search out a higher-order belief that is relevant to my fact-construing processes and imposes a further constraint on where they lead. But higher-order beliefs, for example higher-order beliefs about the consistency of various propositions or scenarios, can also impose constraints on my goal-seeking states, since it will not make sense to set out to realize simultaneously two inconsistent goals. And by the same token they may impose constraints on combinations, not just of beliefs with beliefs, and of desires with desires, but also of beliefs with desires, and of beliefs or desires with actions.

With these points made, I can introduce what I mean by the activity of "deliberation" or "reasoning" or "ratiocination." Deliberation is the enterprise of seeking out higher-order beliefs with a view to imposing further checks on one's fact-construing, goal-seeking, and of course decision-making processes. Not only do we human beings show ourselves to be rational agents, as we seek goals, construe facts, and perform actions in an appropriate fashion. We also often deliberate about what goals we should seek, about how we should construe the facts in the light of which we seek them, and about how therefore we should go about that pursuit: about what opportunities we should exploit, what means we should adopt, and so on. We do this when we try to ensure that we will form suitably constraining higher-order beliefs about the connections between candidate goals and candidate facts.

The fact that we human beings reason or deliberate in this sense – the fact, in Sen's (2002: 40) language, that we conduct "reasoned scrutiny" of our beliefs and desires – means that not only can we be moved in the manner of unreasoning animals by goal-seeking and fact-construing states, such as the belief that p or the desire that q. We can also reflect on the fact, as we believe it to be, that p, asking if this is indeed something we should believe. And we can reflect on the goal we seek, that q, asking if this is indeed something that we should pursue. We will interrogate the fact believed in the light of other facts that we believe, or other facts that perceptions and the like incline us to believe, or other facts that we are in a position to inform ourselves about; a pressing question, for example, will be whether or not it is consistent with them. We may interrogate the goal on a similar basis, since the facts we believe determine what it makes sense for us to pursue. Or we may interrogate it in the light of other goals that

[3] I abstract here from the crucial question of how we come to form concepts like truth and consistency and the like and how we come to be able to form the sophisticated beliefs mentioned in the text. For a little on this see McGeer and Pettit (2002).

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

also appeal to us; in this case, as in the case of belief, a pressing question will be whether or not it is consistent with such rival aims.

Nor is this all. Apart from drawing on deliberation to interrogate the facts we take to be the case, and the goals we seek, we can ask after what actions or other responses we ought to adopt in virtue of those facts and goals. Not only can we ask after whether they give us a reliable position at which to stand; we can ask after where they would lead us, whether in espousing further facts or goals, or in resorting to action. We may be rationally led in the manner of non-human animals, for example, to perform a given action as a result of taking the facts to be thus and so and treating such and such as a goal. But we can also reason or deliberate our way to that action – we can reinforce our rational inclination with a deliberative endorsement – by arguing that the facts, as we take them to be, are thus and so, the goals such and such, and that this makes one or another option the course of action to take; it provides support for that response.

The first possible explanation of why Sen thinks that the belief-desire schema is overly restrictive could be that he thinks deliberation gives the lie to that schema. It does not represent a means, as in the picture I have just sketched, whereby we human beings might hope to discipline the process of belief-formation, desire-formation, and action-selection, using higher-order beliefs as a source of extra checks on that process. Rather, it represents a whole other enterprise: the "disciplined use of reasoning and reasoned scrutiny" (Sen 2002: 19). And this is an enterprise, so the explanation would go, that transcends entirely the regime described in folk psychology and rational choice theory. Let people instantiate that regime in an act of decision-making, mechanically forming and acting on beliefs and desires, and they cannot be deliberating their way to action. Let them deliberate their way to action and they cannot be instantiating the belief-desire regime; they must be operating under a distinct pilot.

But while it is just possible to think that Sen endorses the no-deliberation view of folk psychology, I do not think that this is at all plausible. The thesis is patently false, by my lights, and Sen nowhere shows signs of endorsing it. He never suggests that resort to deliberation and reasoning is inconsistent with our continuing to behave as rational creatures: that is, creatures "whose behavior is regular enough to allow it to be seen as maximizing" (Sen 2002: 30). On the contrary, he envisages deliberation or rational scrutiny as serving often to shape the operation of rational decision-making without disturbing the basic belief-desire regime. The picture in goal-modifying commitment, for example, is that reflection on the needs of others, or something of the kind, can cause us to change the goals we pursue. Such commitment affects how we rationally maximize, it does not undermine maximization itself.

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

2.2 The selfish-deliberation account

Is there any other reason, then, why Sen might be led to countenance the problematic, goal-displacing attitude? A second possibility is worth registering for its inherent interest though, again, I do not think it is likely to be the consideration that moves Sen. The idea here is that the only sort of deliberation that can lead to action, under the belief-desire picture, is deliberation of a self-serving variety. The proposal is not that deliberation has regard only to the agent's private welfare, whether or not this is sensitive to the welfare of others. Rather it is that the very logic of deliberation in goal-seeking or desire-driven creatures ensures that it has a certain self-serving character. Did Sen accept this sort of view, then it too would explain why he might want to insist that something more is often possible: viz., a sort of committed action that escapes the confines of the belief-desire schema.

Suppose that my holding by a certain desire, say that q, makes it rational for me, given the beliefs I hold, to form a further desire or perform a certain action: say to desire that s or to perform B, where each represents an essential means of satisfying the desire that q. And now imagine that I am reflecting on whether there is a deliberative reason why I should hold by that extra desire or perform that particular action. Should I think "I desire that q; so therefore I should desire that s. I desire that q; so therefore I should perform B"? Or will that reasoning leave me with the following question. "Fine, but should I desire that q? Fine, but does this really give me a reason for desiring that s or for performing B?"

I think it is clear that the formula offered will indeed leave me with that question. For we all allow that our desires may not be well formed – any more than our beliefs – and that we may not always have a reason for responding as they require (Broome 2004). Some desires we naturally regard as pathological, others as the products of a weak will, others as due to a lack of imagination or memory, and so on; pathologies of desire abound. This being so, we cannot think that the proper ratiocinative endorsement for acting on a given desire should simply start from the existence of that desire, putting it into the foreground of deliberation, as if it were something sacred and beyond question.

But if we shouldn't deliberate from the fact of what we happen to desire, what should we take as our point of deliberative departure in decision-making (Pettit and Smith 1990)? Presumably, we should enrich the base of deliberation to stipulate that it would be good or desirable or appropriate or whatever to bring about what is desired; holding that this would be good is precisely a way of insisting that the desire is not pathological or wayward in the fashion contemplated. Thus the form of the deliberation that will reasonably take me from what I desire to what else I should desire, or to what I should do, will be: "It would be desirable

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

to satisfy the desire that q; so therefore I should desire that s. It would be desirable to satisfy the desire that q; so therefore I should perform B."

So far so good. But now it is time to notice a clear ambiguity in the formula just offered. What does it mean to think that it would be desirable to satisfy a certain desire? There are three distinctively different readings. An objective reading would say: it means that it would be desirable to fulfill the desire, making the world conform to it, i.e., bringing it about that q. A subjective reading would say: it means that it would be desirable to relieve the desire that q, removing and having the pleasure of satiating the itch that it constitutes, whether or not this means bringing it about that q. And a conjunctive reading would say: it means that it would be desirable to relieve-and-fulfill the desire, with both elements present.

The selfish-deliberation thesis, which I can at last introduce, holds that every form of practical deliberation that leads from desire or preference to action always involves one or other of the relief-formulas: that associated with the subjective reading or that associated with the conjunctive. The idea is that when agents who continue to operate in the belief-desire mould deliberate their way to action – or to the formation of an instrumental desire – they reason from the recognition of having this or that desire within themselves and from the desirability of relieving (and perhaps also fulfilling) this desire.

Why call this a selfish-deliberation thesis? Well, consider how the process envisaged contrasts with how people will reason if they start from the objective reading mentioned earlier. Under that reading, they will ground their deliberation in the fact that it would be desirable to fulfill a desire that q. This is equivalent to grounding their deliberation, more simply, in the fact that it would be desirable to bring it about that q. The formulations are equivalent, because the fact that it would be desirable to fulfill a certain desire – as distinct from the fact that it would be desirable to relieve a certain desire – does not entail that one currently has that desire.[4] Thus people who deliberate in the manner of the objective reading will argue from the desirability of the envisaged state of affairs – that q – without any essential reference to their own state of desire. They will represent that scenario as worth trying to bring about, quite independently of the fact that bringing it about would answer to a desire they themselves happen to harbor.

[4] I assume here that when I think that it would be desirable to fulfill the desire that q, I am thinking: it would be desirable to fulfill any desire that q, not that it would be desirable to fulfill this actual desire that q. The latter formula would suggest, counterintuitively, that what makes it desirable to bring about the fulfilling scenario depends on the existence of the particular desire fulfilled. When I say in the text that the selfish-deliberation model presupposes that one of the relief formulas is relevant, I abstract also at that point from this counterintuitive possibility; it too would make deliberation "selfish."

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

The selfish-deliberation thesis suggests, in contrast to this picture, that all desire-based deliberation represents a prospect as desirable only if it promises to give the agent a distinctive sort of return: to relieve the desire or desires that he or she has for it. The thesis treats all desires as if they were urges or yens, hankerings or yearnings, or pangs: that is to say, itches that exist in consciousness and that naturally demand to be fulfilled and thereby relieved, with all the pleasure that relief would bring (Pettit and Smith 1990).

If Sen thought that all deliberation within the belief-desire schema – all deliberation that is consistent with the rational-choice representation of human beings – had to be selfish in this manner, then we could make immediate sense of why he should think that there are ways of acting that escape the shackles of that schema. There certainly are forms of deliberation – deliberation based on the objective reading of the desire formula, I would say – that are not selfish. And if the only way of countenancing these were to say that human beings sometimes act in committed ways that are not representable within the belief-desire schema, then that might be a reasonable thing to say.

But though adherence to the selfish-deliberation thesis, like adherence to the no-deliberation thesis, would explain Sen's line on goal-displacing commitment, there is little or no evidence that he endorses it. True, the economic way of positing preference-satisfaction as the supreme good in social policy suggests that people are assumed to seek the relief of their preferences, not just their fulfillment, and it may seem to presuppose the selfish-deliberation thesis. But that mode of thought is one that Sen above all others has been vociferous in questioning (Sen 1982). It is very unlikely that he himself should be moved by it.

2.3 The integrated-deliberation account

The no-deliberation account of Sen's line says that so far as agents remain in the belief-desire mould – so far, therefore, as they continue to be representable within rational choice theory – they cannot deliberate at all; hence the acknowledgment of the presence of deliberation requires denying that that schema applies to all human decision-making. The selfish-deliberation account of the line holds that so far as agents remain in that belief-desire mould, they cannot deliberate in a non-selfish way; hence the recognition of non-selfish deliberation requires denying, once again, that the schema applies to all forms of action. The integrated-deliberation account of Sen's line, to which I now turn, maintains that so far as agents remain in the belief-desire mould, and continue to be representable in rational choice theory, they cannot deliberate on the basis of reasons other than those that derive from standing goals that form an integrated system; hence the

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

recognition that people are not just the servants of such standing goals also requires denying that the schema governs everything they do.

Under the folk-psychological picture, every agent who performs an intentional action targets some goal, adjusting his or her behavior with a view to realizing that goal. I assumed in presenting that picture earlier that the goal which is targeted in action may be just about any state of affairs on which the agent might hope to have an influence. Putting the matter otherwise, I assumed that for all that was said, there is no limit on the things that an agent may primitively desire and no limit on the ways in which such desires can come and go. Consistently with the bare bones of folk psychology, the agent may mutate from moment to moment in respect of the primitive desires conceived and acted on.[5]

The integrated-deliberation thesis suggests that the assumption I have been making is mistaken. The idea is that the belief-desire schema, and in particular the rational choice theory that explicates it, posits an image of relatively stable, integrated agents who each come to choice, equipped with a set of dispositions to bring this, that or the other about, and so with a set of corresponding, standing goals. So far as agents act in fidelity to that schema and that theory, so the thesis goes, they will act in the service of those goals. The goals may be entirely benevolent or even self-sacrificing, of course, but they will bear the unmistakeable mark of being goals that belong to the agent involved: goals that have been integrated into the agent's psychology and are truly his or her property.

That the goals envisaged in the thesis are of a standing character does not mean that they have to be long-standing in time. Sen is quite happy to say that goal-modifying commitment can make for a change in someone's goals, leading immediately to action that is intended to advance those newly minted aims. The integrated-deliberation thesis can fit perfectly well with this. What it says is merely that for any goal that one pursues in conformity with the belief-desire schema, the goal has to have been internalized in a distinct episode, however closely related in time; it has to have been "incorporated within one's own goals" as an integrated part (Sen 2002: 35).

The integrated-deliberation thesis will be vacuous, of course, if the internalization and integration of a goal amounts to nothing more than the act of tracking that goal in action. But if it is meant to refer to a distinct psychological episode, one perhaps with a phenomenology of its own, then the thesis will be quite substantive. It will suggest that so far as people conform to the belief-desire schema, they will only act for the furtherance

[5] Continuity will be required only so far as it follows from the sort of continuity that unchanging evidence and information imposes on the agent's beliefs; this will not constrain primitive desires, since they are conceived as nothing more or less than desires that are insensitive to the particular beliefs of the agent (Pettit forthcoming).

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

of goals that have been lodged firmly, if recently, within their psychological make-up. It will imply that so far as human beings register and respond to novel demands – demands unconnected with pre-existing goals – they will transcend the bounds of folk psychology and rational choice theory.

It should be clear that if Sen endorsed the integrated-deliberation thesis then this would explain why he wants to deny that the belief-desire schema is comprehensive. The cases he invokes in relation to commitment typically involve people's registering the goals of others and, without any incorporation or integration of those goals within their own goal-system (Sen 2002: 35), acting with a view to furthering the alien goals, or to furthering some compromise between those goals and their own. Adherence to the integrated-deliberation thesis would make perfect sense of why he claims that in such instances people do not act for the realization of their goals but violate self-goal choice.

It is fairly clear that Sen does not espouse either the no-deliberation thesis or the selfish-deliberation thesis and, given the dearth of alternative explanations for his line about goal-displacing commitment, that already makes a certain case for the thought that he must espouse the thesis of integrated-deliberation. The idea would be that by his lights rational choice theory, and the belief-desire psychology it explicates, presuppose an agent who moves from situation to situation with a pregiven set of goals, modifying those goals only occasionally. Hence, so the idea goes, Sen finds the approach incapable of making room for the extreme of commitment that involves people, not in adjusting their behavior in the light of their own standing goals, but adjusting it in the light of the standing goals of others. According to this interpretation he believes that the way human beings respond to reasons is not limited in the manner allegedly presupposed by the belief-desire schema; he sees goal-displacing commitment as a perfect example of how they can transcend the boundaries of that approach.

I do not myself think that the integrated-deliberation thesis is sound. Neither the belief-desire schema in itself, nor the rational choice theory that seeks to explicate it, has to suppose that the goals which agents try to advance are goals internalized or integrated in any substantive sense. As indicated above, the approach says nothing about how far people may vary in their primitive desires from moment to moment and from context to context.

But if the integrated-deliberation thesis is false, can we ascribe it with any sense of assurance to Sen? I think that perhaps we can. The economic and social-scientific application of rational choice theory naturally assumes that in principle we can fix relatively stable utility and probability functions on individuals; without that assumption the enterprise of explaining and predicting their behavior would be severely jeopardized. Perhaps it is not entirely implausible that the habit of thinking in this methodologically

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

sensitized way should have led Sen to represent rational choice theory as involving an assumption that agents act on integrated goals that generally remain in place from occasion to occasion. This may be precisely what he has in mind when he speaks of the assumption of self-goal choice.

What moves Sen to deny that rational choice theory offers a comprehensive depiction of agency may not be anything so esoteric, then, as the belief that satisfying rational choice is inconsistent with practicing deliberation or that it is consistent only with conducting selfish deliberation. It is perfectly intelligible why Sen might think of the theory as recognising standing goals only. And if that is how he does think then it should be no surprise that his honesty in recognizing the range of people's sensitivity to others would lead him to say that rational choice theory – and so, in effect, belief-desire psychology – does not capture the whole truth about how people deliberate. It should be no surprise that he comes to recognize and celebrate a variety of commitment that he takes to be undreamt of in rational choice circles.

3. CONCLUSION

We saw in the first part of the paper that Sen takes an unusual line in arguing, contrary to some core assumptions of rational choice and common sense, that people sometimes act without acting on goals of their own; they act out of a form of goal-displacing commitment. The exploration of why one might take such a view has led us to identify three theses, any one of which would be sufficient to support Sen's line. I have argued that all three theses are false, holding as I do that Sen is mistaken to think that there is a form of attitude that cannot be accommodated in the belief-desire way of thinking. But it is still interesting to see the grounds on which such theses might be maintained. And it is interesting, in particular, to see that Sen is most likely to have been influenced by the weakest of these three views: the thesis that rational choice only makes room for the form of deliberation in which one argues on the basis of one's pre-existing, integrated goals.

The integrated-deliberation account of why Sen should say what he does about commitment has one merit that I should mention in conclusion. It is a charitable account that does not find anything deeply amiss in his ideas. Consistently with that account, we might even reduce the charge against him to one of terminological infelicity. A perfectly good way of expressing the substance of the account is to say that for Sen the word "goal" means "integrated goal" and the name "rational choice theory" refers to a theory as to what it is rational for agents to do in the light of goals in that sense: that perfectly reasonable sense. This forces Sen into countenancing a range of committed behavior in which notions of goal

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press

and goal-based choice no longer apply, and it leads him into some quite romantic accounts of our capacity on this front. But once the differences of terminology are taken into account, this divergence from more standard models need not make for any great problem. It will count as an indication of idiosyncracy, not a sign of intellectual oversight.[6]

REFERENCES

Broome, J. 2004. Reasons. In *Essays in honour of Joseph Raz*, ed. J. Wallace, M. Smith, S. Scheffler, and P. Pettit. Oxford University Press

McGeer, V. and P. Pettit. 2002. The self-regulating mind. *Language and Communication* 22: 281–99

Pettit, P. 1991. Decision theory and folk psychology. In *Essays in the foundations of decision theory*, ed. M. Bacharach and S. Hurley. Blackwell. Reprinted in P. Pettit (2002)

Pettit, P. 2002. *Rules, reasons, and norms: selected essays*. Oxford University Press

Pettit, P. 2003. Groups with minds of their own. In *Socializing metaphysics*, ed. F. Schmitt. Rowan and Littlefield

Pettit, P. 2005. On rule-following, folk psychology, and the economy of esteem: a reply to Boghossian, Dreier and Smith. Contribution to Symposium on P. Pettit, Rules, reasons, and norms. *Philosophical Studies*

Pettit, P. and M. Smith 1990. Backgrounding desire. *Philosophical Review* 99:565–92; reprinted in F. Jackson, P. Pettit, and M. Smith, eds. 2004. *Mind, morality and explanation*. Oxford University Press

Sen, A. 1982. *Choice, welfare and measurement*. Blackwell

Sen, A. 2002. *Rationality and freedom*. Harvard University Press

[6] This paper was transformed by a searching discussion in the St. Gallen conference on rationality and commitment, May 2004. I am very grateful to the participants there.

https://doi.org/10.1017/S0266267104000367 Published online by Cambridge University Press