Preference, Deliberation and Satisfaction

PHILIP PETTIT

In his famous lecture on 'The Concept of Preference' Amartya Sen (1982) opened up the topic of preference and preference-satisfaction to critical, philosophical debate. He pointed out that preference in the sense in which choice reveals one's preference need not be preference in the sense in which people are personally better off for having their preferences satisfied. And on the basis of that observation he built a powerful critique of some common assumptions in welfare economics.

I endorse Sen's observation and critique and I think that, suitably recast, they can be nicely situated within a broader picture of preference and deliberation that Michael Smith and I developed elsewhere (Pettit and Smith 1990). This paper is an attempt to do just that, sketching an overall picture of the nature of preference, the nature of deliberation, and the way they interact around the idea of preference-satisfaction.

But the paper is not just an attempt to keep the books on these topics; there is also a bottom line. That line is that preference-satisfaction should not normally figure as a deliberative concern. When individuals deliberate about what they ought individually to do, they should not normally focus on what will bring them most preference-satisfaction. And when authorities or commentators deliberate about what good government ought to try to do for its people, they should not normally search for what will maximize the overall preference-satisfaction of people in the community.

The paper is in three sections. The first sketches an overall view of preference, arguing—contrary to Sen, as it happens—that there is one concept of preference, not many.[1] The second outlines the view of deliberation that derives from my joint work with Michael Smith. And the third looks at how preference and deliberation interact and at the place of the idea of preference-satisfaction in this interaction.

[1] Here and in other aspects of that discussion I have been influenced by hearing a presentation of Dan Hausman's at a conference in St Gallen, May 2004.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

Philip Pettit

1. Preference

The first thing to say about the notion of preference is that unlike that of desire, preference is always a preference for one thing rather than another; it always involves a ranking of alternatives. Thus it makes no sense to ask someone whether they prefer X; the only sensible question will be whether they prefer X to Y, X to Z, or whatever. Preference, as we can put it, is inherently comparative.

This introduces straight away a complexity that is often ignored in economics, though it has recently attracted attention among philosophers (Hurley 1989; Broome 1991; Pettit 1991). This is that before we can tell what someone prefers amongst various alternatives, we have to be clear about what exactly those alternatives are. In particular, we have to be clear about how they are individuated, and whether two superficially similar alternatives that appear in difference choice contexts really remain the same option.

Consider in this connection a case where over time you are offered three choices, in each of which another person will get what you leave over. First you are offered a large apple or an orange. Next you are offered an orange or a small apple. And finally you are offered a large apple or a small apple. And suppose that you display a preference for the large apple over the orange, the orange over the small apple, and—surprisingly—the small apple over the large. Does that mean that your preferences are cyclical and, intuitively, irrational? Not necessarily: not if the alternative of taking a large apple and leaving an orange for the other person is a different alternative from taking a large apple and leaving a small apple. And of course those are intuitively different options. The one is perfectly polite, the other downright rude.

In what follows, I shall abstract from the issue of how the alternatives between which people have preferences are to be characterized and individuated. I shall assume that that does not make for a serious difficulty. I have argued elsewhere for a particular resolution of the difficulty but I shall not build particularly on that account (Pettit 1991). So on now to the main topic.

There are three broadly different ways in which we might conceptualize preference, assuming that there is one single concept of preference involved in the way we talk in everyday life and in the manner in which economists conduct their discussions. These different analyses correspond with broader styles of analysis that have been important over the last fifty years or so in the philosophy

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

of mind. The first is a behavioral account of preference, the second a dispositional account and the third a functional one.

The behavioral notion

The behavioral account is associated broadly with the approach described in economics as the revealed-preference theory—see Samuelson (1938) and Little (1950)—though I am not convinced that all revealed-preference theorists would have endorsed it; some, I think, may have aligned themselves with what I go on to describe as the dispositional analysis instead. According to the behavioral account, there is no content to saying that someone prefers one alternative to another over and beyond the claim that he or she chooses that alternative rather than the other. A preference for an alternative is nothing other than what is actually revealed in the choice of the alternative.

This is an extraordinary theory. It means that short of being revealed in choice, there is no preference for anything, so that we cannot say that someone is led to make this or that choice as a result of their preferences and we cannot even say that someone's preference is frustrated by not being able to make a corresponding choice. Equally, we cannot say that people have preferences over matters between which they are unable to choose, whether for the contingent reason that they are not offered the choice or for the deeper reason that the alternatives in question—say, between the world being as it is and the world being dramatically different in some way—are never going to be presented as matters of choice to anyone.

Revealed-preference theory may have lent itself to elegant axiomatization and mathematical development, then, but it looks philosophically very strange. It appears to deny the reality of preference in the accepted sense, rather than giving an account of what that reality involves. In my view it is nothing short of an eliminativist or error theory of preference.

The dispositional notion

The most obvious alternative to the strict behavioral approach retains the tight connection between preference and choice. According to this account, to say that someone prefers one alternative to another is to say that they are disposed, should they

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

be given a choice between those alternatives, to choose the first rather than the second. What does that disposition consist in? The natural way to think of it will be as a categorical state of the agent, or as something grounded in such a categorical state. Thinking of it this way, we can say that when a person chooses the preferred alternative, then the choice is causally explained by the presence of that state within them.[2]

The dispositional analysis gets over the more obvious difficulties with the behavioral. It makes sense of the idea that people are caused to make their choices by the preferences they hold, as it does of the idea that people can have a preference frustrated. And it equally makes sense of the idea that people might have a preference between alternatives that will never be available as matters of choice. I can prefer a perfectly just world, for example, to the world as it is, for were I to be given a choice between those alternatives—impossibly, as it happens—then I would choose the world that is just.[3]

The functional notion

The dispositional account, however, looks to be less general than it ought intuitively to be. It focuses on the connection between preference and choice and makes that connection into something definitional or constitutive; nothing is to count as a preference for X over Y unless it disposes the agent to choose X over Y. But there are connections that we firmly expect a preference to have with other attitudes, and not just with choice, and there is good reason to treat these also as constitutive of preference.

[2] The alternative would be to represent the preference as a bare disposition, which would bring the account back in line with revealed-preference theory. For on the bare-disposition account, there need be no categorical difference between two agents who differ in a choice-disposition and so in some preference.

[3] Can the account make sense of what it might mean to say that I prefer one alternative to another, when the alternatives cannot logically be presented as matters of choice: one might be the alternative of taking a small apple rather than an orange and the other the alternative of taking a small apple rather than a large apple? This issue is raised in (Broome 1991). For an attractive response see (Dreier 1996). His line is that we can have preferences over options, where we abstract from properties that would make the options incomparable.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

The most obvious example of other such connections are the connections between preferences themselves. Suppose that I prefer X to Y, that I am indifferent between Z and not-Z, and that I am offered a choice between having X-and-Z or Y-and-not-Z. Will I prefer X-and-Z over the alternative offered? I must certainly be expected to do so, unless some special, perturbing factors get in the way. Absent those factors—absent temporary insanity or blindness or passion or whatever—a failure to hold by that derived preference would raise a serious question as to whether I really prefer X to Y, or am really indifferent about whether Z or not-Z.

Or consider the sort of connection that holds between preferences and beliefs. Suppose that I am presented with two alternatives, A and B. And imagine that, while the alternatives otherwise leave me indifferent, I believe that A has some desirable property that B lacks; it does not matter how exactly we analyse the notion of desirability or, for that matter, the notion of believing that a feature is desirable. Will I be expected to prefer A to B? Of course I will. Did I not form that preference then, absent some special obtruding factors, it would seem that I cannot be otherwise indifferent between A and B or that I do not really think that A has a desirable property that B lacks.

These observations suggest that we should conceptualise preference so that the connections between a preference and other states are given the same definitional prominence that is given, under the dispositional account, to the connection between a preference and choice. After all, the failure of those connections, like the failure of the connection with choice, would raise a question as to whether there is a preference present at all.

These observations, if we go with them, take us toward a functional analysis of preference. According to such an analysis, to say that someone prefers one alternative to another is to say that they are in a state such that, in the absence of perturbing conditions, that state will dispose them to choose the first rather than the second, and will connect in such and such a manner with other preferences and other states of mind. Preferences, roughly speaking, are not just any sorts of dispositions to choice; they are dispositions that are collaterally sensitive to a variety of other states.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

How exactly will a preference have to connect with other states? Without going into detail, it is worth observing that decision theory can give us a lead on this question.[4]

There are certain intuitive connections that are important to the notion of preference—connections like those that link them with judgments of desirability, for example—that are not registered in decision theory (Pettit 1991). But the theory does map a range of connections of a kind that certainly are relevant, suggesting that for a given state in the agent to count as a preference, and as a preference with a specific content, it must relate to other states in a certain pattern, at least in the absence of perturbing factors. Thus it stipulates that for one state to be a preference for X over Y, and for another state to count as indifference between Z and not-Z, the states must connect in such a way that they give rise to a preference for X-and-Z over Y-and-not-Z. Decision theory consists in a set of axioms that dictate a range of such functional connections that *bona fide* preferences must satisfy.

The connections marked in decision theory include connections to preferences, not just over particular alternatives, simple or compound, but preferences over probabilistic gambles involving those alternatives: preferences over gambles assigning such and such a probability to one alternative, and such and such a probability to another. It turns out that by doing this, it makes it possible for each alternative over which an agent registers relevant preferences to construct an index of how relatively intensely the agent prefers that alternative; the scale whereby those intensities are determined is known as the agent's utility function (Ramsey 1990). The degree of preference that is thereby determined for an alternative can be represented as corresponding to the person's desire for that option; it will attach to each alternative within the agent's preference-ordering and can be attached without mention of any explicit alternative.

I have argued that we ought to conceptualize preference so that any state that is to count as a state of preferring one alternative to another should connect in certain ways, at least in the absence of

[4] There is one important complexity to note. This is that a full functional analysis will need to provide an analysis, simultaneously, of what is involved in someone's having each of the preferences they display, not just a single one, and perhaps each of the other connected attitudes as well. It will have to be holistic in the sense of conforming to the familiar Ramsey-Carnap-Lewis framework for functional analysis (Lewis 1983, Essay 6). Decision theory might be recruited to this holistic task.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

perturbing factors, with choice, with other preferences, with beliefs of various sorts, and so on. Decision theory gives us a good lead on the connections that anything deserving the name of 'preference' should be expected to satisfy, though not a lead on all plausible connections; I mentioned as an example the connection to judgments of desirability. I do not mean to go further into positive detail about the connections that are important to preferences but there are three negative remarks that I should certainly make. They are independently intuitive and they combine to provide a workably specific notion of preference.

The three remarks are that preferences in general should not be expected to have the connections associated with matters of taste, feeling and self-interest. As a rule,

- they are not disconnected attitudes like tastes;
- they do not have any phenomenal or felt quality; and
- they do not spring from self-interested desires.

These remarks are important because there are models under which preferences are nothing but tastes, preferences are conscious, qualitative phenomena, or preferences are invariably self-interested.

To hold that preferences are tastes is to suggest that they are brute states in which one finds oneself, as one finds oneself with a taste for dark beer or bright clothes or the smell of garlic. In particular, it is to suggest that they are exogenous to decision-making and are not themselves up for adjudication or revision. There is no debating about tastes; *de gustibus non disputandum*. There is no debating about tastes and, as the other cliché has it, no accounting for tastes. But what is true of tastes in these regards is certainly not true of preferences in the functional sense in play here. Preferences in general are susceptible to deliberative connections with a variety of factors—more on this in the next section—and do not have the insulated, unmoveable character of tastes.

One reason why people might think that preferences are like tastes is that they think of preferences, more specifically still, as phenomena with their own qualitative feel. They are taken to be the sorts of conscious inclinations that we describe as yens and hankerings, urges and impulses, cravings and passions and itches. All of these attitudes have a phenomenal quality in the sense that there is something it is like to have them. And because they represent such a salient if not common aspect of decision-making, they are easily taken as the basis for a model of preferences. But any such model would be quite misleading. Understood in the

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

functional sense, it is quite clear that most preferences—most collaterally sensitive dispositions to choice—do not have a phenomenal side.

The final remark I want to make about preferences is that as there is no reason to take them as tastes or as itches, there is no reason to think that they necessarily connect—as of course tastes and itches might be thought to connect—with the self-interest of the agent. I may have a preference defined over any alternatives, no matter how disconnected from my sense of my own welfare. I may even have a preference for one alternative rather than another, when that alternative promises to do worse by my personal welfare than the other. There is no incoherence, and every plausibility, in the idea of my instantiating and acting on such a detached ranking of options.

This completes my discussion of the notion of preference itself. It is time now to turn to the second topic of deliberation. With that topic covered, we will be able to turn to the interaction between preference and deliberation and to look at its significance for the role of preference-satisfaction.

2. Deliberation

Folk psychology and decision theory

The fundamental tenet of our common sense psychology of human agents is that agency involves acting to realize various goals in a way that is sensible in light of the apparent facts: that is, in a way that adjusts to the facts, as one construes the facts (Jackson and Pettit 1990). Agents seek goals, construe facts, and choose an action that will achieve their goals—or will maximize the chance of their goal being achieved—if the facts answer to how they are construed. For short, people act so as to promote their goals according to their construal of the facts.

This common sense view—this folk psychology—can be just as well expressed in the language of preference, which is exactly what decision theory does. The output of decision-making under this variant is the formation of a preference ordering over the options available in a choice; this then leads directly to choice and action. The inputs are the agent's background preferences over the possibilities that action might affect—the agent's degrees of preference for those scenarios—together with the agent's degrees of confidence or probability that this or that scenario will be realized

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

in the event of this or that action being taken. The agent's goals are the scenarios that attract relatively high degrees of preference and the agent's degrees of probability represent his or her construal of the facts.

Putting the two schemas together, the view shared between folk psychology and decision theory goes, roughly, like this.

- To seek certain goals is to be in corresponding goal-seeking states, described in common sense as desires; these are represented in decision theory by relatively high degrees of preference or utility.
- To construe facts is to be in corresponding fact-construing states, described in common sense as beliefs; these are represented in decision theory by degrees of probability or confidence.
- To seek certain goals according to how one construes the facts is to be caused to act—not by accident but 'in the right way' (Davidson 1980)—by the presence of the relevant complex of belief and desire, probability and preference.
- More particularly, it is to be caused to act so that the agent's desires are promoted according to the agent's beliefs—so that the agent's expected utility is maximized, with the option that attracts the highest degree of preference being selected.

Given the concordance between talk of seeking goals and construing facts, and talk of preferences and probabilities, we can speak indifferently in either idiom. When it comes to situating deliberation in human decision-making, the folk-psychological idiom of goals and facts is easier to work with and this is how I shall mainly write in this section. The issue of how to place deliberation in relation to preference is just the issue of how to place it within the folk-psychological schema of goal-seeking, fact-construing agency: of an agency of belief and desire.

Introducing deliberation

The first thing to notice in approaching the topic of deliberation from this angle is that folk psychology, understood as the affirmation of the goal-seeking, fact-construing nature of agency, may apply in the absence of anything we would naturally describe as deliberation. This comes out in the fact that by most accounts, though not by all (Davidson 1980), the psychology is true of non-human as well as human animals.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

The idea is that many 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 behaviour. They will be rational agents in the sense of conforming to folk psychology and decision theory. Or that will be so, at any rate, in the absence of intuitively perturbing influences, and 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. 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 beliefs about those goals we pursue or might pursue and those facts believe or might believe; 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 ask questions about those properties and relations of goals and facts, with the beliefs we form in response to that interrogation serving as checks on the overall pattern of attitudes that is going to unfold within us (Pettit 1993, Ch 2; McGeer and Pettit 2002).

The exercise whereby we impose such 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 the higher-order level about the candidate fact that p and 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 'p' and 'r' are inconsistent, and so my belief-forming

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

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.[5]

In this example, I find a higher-order truth—that 'p' and 'r' are inconsistent—which is relevant to my fact-construing processes and imposes a further constraint on where they lead. But the higher-order truth recognised in the example could equally have had an impact on my goal-seeking processes; it would presumably have inhibited the simultaneous attempt, for example, to make it the case both that p and that r.

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 truths—truths about consistency, support and the like—with an implicit or explicit view to imposing further checks on one's fact-construing and goal-seeking 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 beliefs in suitably constraining higher-order truths about the properties and relations of candidate goals and candidate facts.

The fact that we human beings reason or deliberate in this sense means that not only can we be moved by goal-seeking and fact-construing states—by the belief that p or the desire that q—in the manner of unreasoning, if rational, animals. 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

[5] 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/S1358246100009498 Published online by Cambridge University Press

makes sense for us to pursue. Or we may interrogate it in the light of other goals that 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.

One final comment. Drawing on deliberation in full explicit mode, as this account suggests, involves asking after certain higher-order matters. But I may be subject to deliberative control even in cases where I do not explicitly deliberate in this sense. Suppose that without explicit deliberation I tend generally to go where such deliberation would lead me and that if I do not—if my habits take me in intuitively the wrong direction—then the 'red lights' generally go on and I am triggered to activate deliberative pilot. Under such a regime, deliberation will 'virtually' control the evolution of my beliefs and desires; it will ride herd on the process, being there as a factor that intervenes only on a need-to-act basis.[6]

The truth-serving, value-serving structure of deliberation

So much by way of introducing deliberation. But what exact form does deliberation take? What are the premises invoked when I deliberate my way to some novel conclusion, whether a conclusion that I should believe such and such, desire so and so, or choose this or that action?

Suppose that my holding by a certain belief, say that p, makes it rational to form a further response: for example, to hold by an

[6] See Pettit (2001), Ch.2.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

entailed belief that r or, given suitable goals, to perform a certain action, A. And now imagine that I am reflecting on whether there is a reason why I should hold that r or perform A. Should I think 'I believe that p; so therefore I should hold that r, or perform A'? Reflection suggests not; or it suggests at least that I should not confine myself to this project (Broome 2004). There will always be a question as to whether I should believe that p. And if it is not the case that I should believe that p, then there may not be any reason to believe that r or to perform A. It may be that I am mistaken or unjustified in believing that p, for example. It may even be that the belief that p is lodged unmoveably within me, despite all the evidence I register against it (Dennett 1979); it may represent a sort of pathology.

How should I deliberate and think, then, if I am to raise the question as to whether there is a reason to believe what 'p' entails or to act as it suggests I should act? Clearly, I should ask whether p; and if I remain convinced on that score I should reason: 'p; so therefore I should believe that r; p, so therefore I should do A'. It will be the fact that p, as I take it to be, that provides a reason for holding by the further belief, or taking the relevant action, not the fact that I believe that p. And this formulation makes that feature salient. An alternative that would do equally well, of course, is: 'It is true that p; so I should believe that r. It is true that p; so I should perform A'. For the fact that it is true that p means, not that I believe it, but that I should believe it; and in this way it serves in the same role as the fact that p. Weaker alternatives that would also serve appropriately, though not with the same force, are 'probably p, so ...' or 'it is probably true that p; so ...'. But this is not the place to get into such detail.

The question that arises now with desire is whether things go in parallel there to how they go with belief. 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. And now imagine that I am reflecting on whether there is 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 leave me without the fullest ratiocinative endorsement available? Will it leave me with the thought: '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 leave me with that question. For as we allow that our beliefs may be false or

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

ill-founded, and that we may not strictly have a reason for responding as they require, so we all allow that our desires may not be well-formed and that equally we may not always have a reason for responding as they require. 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 are a lot more commonplace than pathologies of belief. 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.

What form will the ratiocinative endorsement of desire take? It cannot parallel the example with belief that goes 'p; and so ...'. But it can parallel the variant that invokes the truth that p, or the likelihood that it is true that p. As the truth of something means that I should believe it—that I have a reason for believing it—so the property of a goal that we ascribe when we say it is 'desirable' or 'good' or 'appropriate' or 'valuable' means that I should desire it, that I have at least a defeasible reason for desiring it. Assuming that there is some property deserving to be named by such a term, we can say that in deliberating our way to action we have to take our start, not from the fact that we desire certain goals, but from the fact that, as we see things, those goals are desirable or good or valuable or whatever. This line fits with our ordinary practice and with the long tradition of thinking that the major premise in a practical syllogism should not mention the fact that some state of affairs is desired but rather the fact that it is worthy of being desired (Anscombe 1957).

The picture of deliberation emerging from these considerations is that it is a truth-serving and value-serving enterprise. Deliberation tries to track the true and the valuable, not the believed and the desired, in looking at whether a novel response is well supported. And this is the case whether the response is the formation of a new belief or desire—or indeed a novel intention or policy or the like—or the performance of an action.

So far as the model depicts deliberation as truth-serving, it fits with received wisdom and will raise few questions. But won't it be more controversial in depicting deliberation as value-serving? Won't it be more controversial, in particular, when it assumes that there is a property of goals that deserves the name of 'desirability' or whatever? Truth, it may be said, is a relatively uncontested reality—at least outside of some postsmodernist circles—but desirability or value is inherently questionable.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

The comment to make in response to this worry is that what makes a goal worthy to be desired may be held to consist in any of a variety of features but that almost every philosophical view will countenance some features that play this role; it will acknowledge that there are some value-making properties. That a goal counts as desirable, or at least desirable for me, may be held to require one or more of the following properties, for example:

- it is something I can bring about;
- it coheres with other things I want to bring about;
- it isn't the sort of goal that, once achieved, will seem like nothing (Milgram 1997);
- its attraction doesn't depend on any false beliefs, any failures of reason, any temporary derangements of sentiment, and the like (Smith 1994);
- my pursuit of the goal can be justified to others, serving to further common ends (Pettit 1997), or to further a sectional end that others can endorse (Scanlon 1998).

A complexity with desire

The model of deliberation introduced so far abstracts from a complexity with desire that is important to mention in concluding this discussion of deliberation. While all desires are goal-seeking states, capable of being characterized by their functional role, some desires also have a phenomenal aspect; there is something it is like to have them. Or at least that is a natural way to characterize certain examples. I am thinking of desires like the craving for a cigarette, the yearning for a drink, the ache of loneliness. With such desires we are not just conscious of the states of affairs that they make attractive to us; we are conscious of the states of feeling or inclination in which they themselves consist or by which they are attended. Those states have the presence of a disturbing itch, so that it makes perfect sense to think of endorsing a certain response—going for a smoke, getting something to drink, arranging to meet some friends—on the grounds, at least in part, that this will relieve that itch; this will restore equilibrium.

What one thinks desirable in such a case will not be the state of affairs considered in itself—the smoke or the drink or even the social gathering—or at least not exclusively that state of affairs. What one thinks desirable, at least in part, will be that state of affairs considered as a means of relieving the phenomenal

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

desire—as a means of ensuring that the craving or yearning or ache has gone, with the pleasure associated with getting rid of it in that way as distinct from getting rid of it by resort to therapy, or treatment, or drugs. The existence of phenomenal as distinct from merely functional desires marks a genuine disanalogy between the case of belief and desire, between fact-construing and goal-seeking states. The disanalogy reveals that there are two sorts of desires. With one, the goal is simply the desired state of affairs in itself; with the other, it is, at least in part, the desired state of affairs in its role as a means of relieving the desire.

The fact that there are two sorts of desire, phenomenal and non-phenomenal, does not undermine the value-serving model of deliberation. According to that model, the reasoned endorsement of a response which a desire makes it rational to form should not invoke the existence of the state of desire but rather the desirability of the objective desired. And the mere existence of a phenomenal desire will not provide a warrant for acting so as to satisfy it. If such action is to be warranted, then it must be the case, not just that I have the desire, but that it is desirable that the desired state of affairs be realized and the desire be thereby relieved. The fact that I have a craving for a smoke will not provide a warrant for smoking except so far as it is desirable, if indeed it is, that the craving be relieved in the distinctive manner of satiated fulfilment.

But though the existence of phenomenal as well as merely functional desires does not undermine the value-serving model of deliberation sketched in this section, it does force us to be careful. It shows that we have to be alert to whether the goal-seeking state on the basis of which someone acts is phenomenal or non-phenomenal in thinking about what form their deliberation will take. This point will prove to be important to the discussion in the next section.[7]

3. Preference meets deliberation

We come finally to the question of how preference and deliberation interact in decision-making and, in particular, of what role the idea

[7] Many of the points made here about phenomenal desire apply more generally to desires such that the agent may wish to be rid of them, whether or not they have a phenomenal aspect: say, unconscious impulses or obsessions. Like phenomenal desires, these will be such that it will only make sense to act on them in the event that they continue to be present; see the next section.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

of preference-satisfaction plays. I will conduct the discussion by looking in turn at three questions. First, what do I target in deliberative decision-making? Second, what makes the target or targets desirable for me? And third, what are the implications for the role of preference-satisfaction?

What do I target?

Asked what I target in deliberative decision-making, I might well answer: the satisfaction of my preferences. This would fit perfectly with everything that we have seen so far. But the various points made in the last section suggest that I should go on immediately to explain exactly what I mean and to guard against some obvious misunderstandings.

What needs to be said in particular is that the satisfaction of preferences may refer either to the realization of those states of affairs that fulfill the preferences or to the relief of the preferences, as we might call it: the removal of the preferences from one's psychology by means of fulfilling them. It is perfectly reasonable to think in the case of phenomenal desires that not only do agents seek to fulfill their preferences, they seek also to be relieved of the phenomenal itch that those preferences constitute—and of course to enjoy the pleasure associated with the relief process.

So what do I target in deliberative decision-making? In the normal, functional case, it should be clear that what I target is the fulfillment of my preferences: the realization of those states of affairs that attract sufficient intensities of preference—sufficient desires—to count for me as goals. In the case of phenomenal desires, however, I may have a double target: the fulfillment of my preferences in the sense just explained plus the relief that fulfillment will bring. Conceivably, however, I might be indifferent to the relief and seek only the fulfillment. Or I might be indifferent to the fulfillment and seek only the relief.[8]

[8] I might also just want to get rid of the desire, whether with or without the pleasure associated with relief. In this latter case I would presumably be happy to have the preference or desire removed by therapy or treatment or drugs, rather than by fulfillment; indeed, for reasons to do with long-term prospects, I might even prefer this mode of removal.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

What makes my target or targets desirable for me?

So much for what I target in deliberative decision-making. But what now makes my target or targets desirable for me? What is it about the fulfillment of my preferences, or the relief of my preferences, that is likely to appeal to me when I am drawn to them? What are the desirability characteristics—the cherished features—that I am liable to register in those goals (Anscombe 1957)?

The main claim I want to make about the desirability characteristic of preference-fulfilment in the normal, functional case is that it is not necessarily egocentric. Perhaps I have to think of what I set out to achieve in ego-relative terms as fulfilling my preference. But the feature that will make the sought-after state of affairs attractive or desirable for me is not necessarily or generally that ego-relative sort of property. It will be for other reasons, and not because the action promises to fulfill my preferences, that I will find it desireable.

Were it important to me that some future state of affairs that I prefer to have realized will serve to fulfill my preference then, plausibly, I will only want to have it realized in the future so long as my preference itself continues into the future (Parfit 1984). But there is nothing even slightly unusual about taking steps for the fulfillment of a normal, functional preference in full awareness that by the time fulfillment comes the preference will have disappeared. I may now act on a preference for committing a fortune I will eventually inherit to the poor, well aware that by the time that I inherit the money the anxieties of age will have taken over and I will no longer have the preference on which I acted earlier. Again, I may now act on a preference for publishing a paper, conscious on the basis of past experience that by the time it appears I will have lost the preference for having it in print. These phenomena would make no sense if what was supposed to make giving away my fortune or publishing my article desirable was the fact that it would fulfill a concurrent preference.[9]

The lesson, I think, is clear. In seeking the fulfillment of a preference, the desirability characteristic of the fulfilling state of

[9] At least these phenomena would not make sense, assuming that the supposed desirability characteristic was that realizing the preferred state of affairs in each case would fulfill a contemporary or concurrent preference. It remains possible in principle that I might be focused on this desirability characteristic instead: that when the preferred scenario comes to be

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

affairs need not be ego-relative and the preference need not be egocentric. I may find that state of affairs desirable for egocentric reasons, of course, such as that it will further my prospects in life. But again I may find it desirable for a variety of other more altruistic or neutral considerations too: that it will help you or some others in this or that manner, that it will make for greater justice in the world, that it will increase the sum of sentient happiness, or whatever.

So much by way of commentary on the regular case of functional desire. But it is worth remarking that there is a contrast here with phenomenal desire. Suppose that I want something, not just because of the inherent appeal of the preference-fulfilling state of affairs—the inherent appeal, if there is any, in scratching or smoking or whatever—but because of the appeal of that action as a way of relieving me of a phenomenal desire: assuaging the itch, satiating the yearning for nicotine. In such a case I will be preferring the state of affairs sought for a characteristically ego-relative property and the preference will be to that extent egocentric. Thus it may make little or no sense for me to arrange to have the preference satisfied at a later time, if I believe that at that later time the preference will already have disappeared. To believe this will be to believe that the fulfilment of the preference—the fulfilment in the future of what will be by then a past preference—cannot have the desirability characteristic of relieving a phenomenal state.

We have seen that while I may be said to pursue the satisfaction of my preferences in deliberative decision-making, this is ambiguous between saying that I pursue fulfillment of the preferences and saying that I pursue relief from the preferences. We have seen that only fulfilment is relevant with normal, functional preferences but that relief is relevant—on its own or alongside fulfilment—with preferences of a phenomenal kind. And we have seen, finally, that whereas preference-fulfillment may be desirable for non-egocentric or egocentric reasons, there is something essentially egocentric about the desirability of preference-relief.

---

realized it will serve at that time to fulfill retrospectively the preference that I now have for its realization. This possibility is so bizarre, however, that we need not delay over it.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

What are the implications for the role of preference-satisfaction?

There are two implications for the role of preference-satisfaction supported by these lines of argument. The first is that it is potentially misleading to frame one's practical decision-making in terms of satisfying one's preferences. And the second is that it would be a serious mistake for policy-makers to think that increasing people's preference-satisfaction is a sensible goal.

To frame one's decision-making in terms of preference-satisfaction would be to deliberate, implicitly or explicitly, along these lines: I have preferences such that the best way of satisfying them is to perform an action, A; it is desirable to satisfy my preferences; so I should perform action A. But this mode of reasoning suggests that what makes the satisfaction of my preferences desirable is not the character of the preference-fulfilling scenario in itself, or least not entirely that. What is also crucial, so the implicature goes, is the fact that the scenario has the ego-relative property of fulfilling my preferences.

This will not generally be an implicature I should endorse. It will be sound in the case where the preferences are phenomenal in nature and their fulfillment is desirable in part for the relief it will bring. But it will not generally carry. Consider the case, for example, where I am persuaded of the value of helping to ameliorate third world poverty by making a financial contribution. What is likely to make that contribution desirable by my lights? It just might be that doing so will relieve a guilt-related preference for making a contribution. But in the more regular sort of case the contribution will be desirable by my lights for reasons unrelated to any such egocentric return. That might appear, for example, in the fact that I am happy to precommit some future earnings to the cause of third world poverty, even as I foresee that I will later come to have different, more self-serving preferences and feel no guilt whatsoever.[10]

Were I always to frame my practical reasoning in terms of preference-satisfaction then there might be a danger of losing the distinction between seeking preference-fulfilment and seeking preference-relief. It might begin to seem that acting on one's preferences always means acting for a sort of personal advantage and that egocentricity is built into the very logic of human

[10] It is just possible in this case of course that I now precommit to contribute in the future because of the relief from guilt that doing this gives me now.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

decision-making. I think it is very important to resist this mistake, if only to guard against a sort of global demoralization about our species.

The mistake is not always avoided, particularly among the economically minded theorists who make the notion of preference-satisfaction central in their accounts of agency. Consider this remark, for example, from Anthony Downs (1957, 37), a classic exponent of the rational choice approach to politics. 'There can be no simple identification of acting for one's greatest benefit with selfishness in the narrow sense because self-denying charity is often a great source of benefits to oneself. Thus our model leaves room for altruism in spite of its basic reliance upon the self-interest axiom'. Downs's idea is not that it pays to be charitable—he is not a homespun philosopher—but that acting for charity, presumably because it is a case of acting for the satisfaction of one's own charitable preferences, will inevitably have a self-interested aspect. He suggests that self-interest will be present as much in altruism, then, as in 'selfishness in the narrow sense'.

So much for the first implication of our considerations, that it is potentially misleading to frame one's practical decision-making in terms of satisfying one's preferences. The second implication is that it would be a serious mistake for policy-makers to think that increasing people's preference-satisfaction is a sensible goal.

The point here is one that Amartya Sen (Sen 1982; Sen 2002) is famous for having emphasized and it fits nicely with our picture of preference and deliberation. Suppose that people were as inescapably egocentric as they would be were every form of practical deliberation directed towards the relief as distinct from the fulfilment of preference. In that case they could each be represented as seeking their own advantage—the satisfaction of their own preferences, as in the relief that that provides. And now consider under that supposition how people in government might reasonably think about the point of the policies they are to introduce. It would make perfect sense for them to act—assuming that their own egocentric concern allows them to act—for achieving the best overall satisfaction of people's individual preferences. If individuals are supposed to be concerned only with their own welfare—the satisfaction, as in the relief, of their own preferences—then it must seem reasonable to think that an agency that acts in their name as a group should be concerned with looking for an arrangement in which their individual preferences are equally satisfied, or in which the net balance of preference-satisfaction is maximized, or something of the kind.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

This, however, will look like an absurd policy-goal once it is recognized that, as we have been arguing, people are often quite non-egocentric in acting on their preferences. Suppose that some people act out of egocentric preferences, and others out of more or less altruistic preferences: say, preferences for the welfare of others, including others of an egocentric bent. Then a government that sought to equalize or maximize preference-satisfaction in the society would be double-counting the egoists. They would be looking after them on two counts: both as objects of their own concern and as objects of concern to the more altruistic.

The point here, like the point of the first implication, is hard to miss once it is spelled out but easy to miss in the absence of some emphasis. I hope that situating the two points within an overall view of what deliberation and preference are and of how they work together will help to make them absolutely inescapable. They are platitudinous in character but they represent platitudes that economic theorizing and policy-making has sometimes proved capable of overlooking.[11]

References

Anscombe, G. E. M. (1957). *Intention*. Oxford, Blackwell.

Broome, J. (1991). *Weighing Goods*. Oxford, Blackwell.

Broome, J. (2004). Reasons. *Essays in Honour of Joseph Raz*. J. Wallace, M. Smith, S. Scheffler and P. Pettit. Oxford, Oxford University Press.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford, Oxford University Press.

Dennett, D. (1979). *Brainstorms*. Brighton, Harvester Press.

Downs, A. (1957). *An Economic Theory of Democracy*. New York, Harper.

Dreier, J. (1996). 'Rational Preference: Decision Theory as a Theory of Practical Rationality.' *Theory and Decision* **40**: 249–76.

Hurley, S. (1989). *Natural Reasons*. New York, Oxford University Press.

Jackson, F. and P. Pettit (1990). 'In Defence of Folk Psychology.' *Philosophical Studies* **57**: 7–30; reprinted in F.Jackson, P.Pettit and M.Smith, 2004, *Mind, Morality and Explanation*, Oxford, Oxford University Press.

Lewis, D. (1983). *Philosophical Papers Vol 1*. Oxford, Oxford University Press.

[11] I am very grateful for the many helpful comments I received when the paper was presented at the Royal Institute of Philosophy Conference in Cambridge, July, 2004, and when it was discussed in a seminar I taught with Adam Elga in Princeton, Fall 2004. Jeremy Butterfield provided a particularly useful set of comments and I am in his debt.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press

Little, I. M. D. (1950). *A Critique of Welfare Economics*. Oxford, Oxford University Press.
McGeer, V. and P. Pettit (2002). 'The Self-regulating Mind.' *Language and Communication* **22**: 281–99.
Milgram, E. (1997). *Practical Induction*. Cambridge, Mass., Harvard University Press.
Parfit, D. (1984). *Reasons and Persons*. Oxford, Oxford University Press.
Pettit, P. (1991). Decision Theory and Folk Psychology. *Essays in the foundations of Decision Theory*. M. Bacharach and S. Hurley. Oxford, Blackwell; reprinted in P.Pettit 2002 *Rules, Reasons, and Norms*, Oxford, Oxford University Press.
Pettit, P. (1993). *The Common Mind: An Essay on Psychology, Society and Politics*, paperback edition 1996, New York, Oxford University Press.
Pettit, P. (1997). A Consequentialist Perspective on Ethics. *Three Methods of Ethics: A Debate*. M. Baron, M. Slote and P. Pettit. Oxford, Blackwell.
Pettit, P. and M. Smith (1990). 'Backgrounding Desire.' *Philosophical Review* **99**: 565–92; reprinted in F.Jackson, P.Pettit and M.Smith, 2004, *Mind, Morality and Explanation*, Oxford, Oxford University Press.
Pettit, P. (2001). A Theory of Freedom, Polity, 2001.
Ramsey, F. P. (1990). *Philosophical Papers*. Cambridge, Cambridge University Press.
Samuelson, P. (1938). 'A Note on the Pure Theory of Consumers' Behavior.' *Economica* **N.S. 5**: 61–71.
Scanlon, T. M. (1998). *What We Owe To Each Other*. Cambridge, Mass., Harvard University Press.
Sen, A. (1982). *Choice, Welfare and Measurement*. Oxford, Blackwell.
Sen, A. (2002). *Rationality and Freedom*. Cambridge, Mass., Harvard University Press.
Smith, M. (1994). *The Moral Problem*. Oxford, Blackwell.

https://doi.org/10.1017/S1358246100009498 Published online by Cambridge University Press