

Reply to Critics

ALLAN GIBBARD

University of Michigan, Ann Arbor

I thank all my critics in this symposium for these marvelous commentaries. The misinterpretations are few, and those few concern the most subtle aspects of my view. The commentaries get to the heart of a number of the most important issues that my theory of normative concepts raises. They show at the very least that I need to clarify what I say in the book, and at some points they say things that demand rethinking. I'll begin with Dreier and some of the logical issues at the heart of the theory, and then move to Bratman on the rich features of plans and intentions and the possibility of two alternatives' being incomparable. These issues all arise for ideal planners, beings who make no conceptual mistakes in thinking what to do and who do what their plans say to do. I then move to Scanlon on *acrasia*, and ask what I should say if I agree that clear cases can arise where the will doesn't follow a person's *ought* conclusions. I end with Blackburn and Sinclair on how my view contrasts with normative realism.

My strategy of inquiry in the book was to start with an ideal thinker-planner. What concepts are we to credit her with using? Only later did I turn to our own concepts, and ask whether we share the concepts an ideal thinker-planner would deploy. Of course a treatment of the ideal planner must leave much unsaid about us who are far from ideal. We ourselves are inconsistent and fail to see many of the consequences of the things we believe. We fail to keep track of the consequences we do see. We conclude what to do and then don't do it. Still, if we want to understand the logic of our concepts, I thought, we'd best first look to thinkers who are consistent, and then ask later how to interpret our own muddles. The features of a concept, after all, are revealed in its consistent employment. That leaves the problem of tying these features to us, with our complex and messy psyches.

Ramsey and his successors adopted the kind of approach I had in mind for decision theory.¹ He set out to explain the concept of probability, the concept we express with terms like "maybe", "probably", and "almost certainly". His

¹ Ramsey, "Truth and Probability" (1931). See also especially Savage, *Foundations of Statistics* (1954), and Hammond, "Consequentialist Foundations" (1988).

method was to ask what consistency would amount to in preferences among risky prospects. He discovered that from a consistent chooser, we can read off a concept that fits the classic laws of probability. This was the basis of his account of our own concept of probability. Far more needs to be said about such an approach than I know how to say, but with Ramsey the approach was immensely fruitful.

As I address the points the critics raise, then, it will help if I distinguish the two stages of the inquiry: First, the ideally consistent thinker-planner. She, my hypothesis is, gives normative concepts their logic. Only next come we who deploy these concepts imperfectly. An adequate treatment of normative concepts as we use them, I have no doubt, requires a far richer psychology than any I offer. It is difficult to sort out, though, which questions about normative concepts arise at which of these stages. Take elementary logic as a parallel: We can try using a thinker's states of mind to characterize concepts of entailment and the like. If, though, we insisted at the outset on full psychological realism, we'd get a mess. In my earlier book *Wise Choices, Apt Feelings* (1990), I did speculate on a richer psychology, and tried to account for such things as a distinction between a demand of rationality and an existential commitment. In *Thinking How to Live*, in contrast, I paid little attention to any non-ideal psychology. Rather, I tried to delineate narrow questions on which the ideal thinker-planner could provide lessons. Some of the critics' qualms and objections concern whether my treatment works even for ideally consistent planners. Others concern whether, even if it does, the treatment applies to the concepts that we ourselves employ.

Plans and Beliefs

Begin, then, with a thinker-planner who seems in no way inconsistent. This planner, I argued, would have a concept of (as I put it) being the *thing to do*. Sherlock Holmes worries about whether it is now time to pack his bags. To conclude that packing is now the "thing to do", I started out saying, is to decide to pack. This dictum, though, could only be provisional. If the ideal planner concludes that packing is now the thing to do, he does forthwith pack. If he packs, though, he might still not think packing to be *the* thing to do. He might be indifferent between packing and some alternative, such as fleeing without packing, and plump for packing out of indifference. In that case, as we can say, he *permits* himself to pack, but also permits fleeing. For an ideally consistent planner, packing shows that he permits packing, but it doesn't go the other way around: he can permit immediate fleeing and not flee. What one picks doesn't fully reveal which alternatives one permits and which one rejects.

How much of normative concepts, then, can we glean from consistent planning? That will depend on what counts as "planning". In the book I

spoke of “contingency plans”, and argued that any planner is committed to concepts that act much like certain of our own ought concepts. If a “plan”, though, just consists in picking an alternative for each contingency, the “planner”, the picker of alternatives, might not have anything close to genuine normative concepts. The critics all stress this. I myself eventually gave a different meaning to the term ‘plan’. A plan as I ended up using the term allows or forbids each alternative, allowing at least one alternative for each contingency. Dreier adopts a different usage, and so to distinguish what he calls a “plan” from what I use the term to mean, I’ll call a “plan” in Dreier’s sense a “strategy”.² A *strategy*, I’ll say, picks out a single alternative for each contingency, whereas a *plan*, as I use the term, permits or rules out each alternative, permitting at least one.

“Planning” in my sense differs from planning in the ordinary sense in two further important ways. First, I allow what one will in fact do to be part of the situation planned for. The binge alcoholic can plan, I say, to abstain on Saturday night, where part of what he knows is that in fact he will drink. Second, I allow for contingency planning that is entirely hypothetical, for contingencies one knows one won’t face.

I find no serious problem with either of these two departures. Hypothetical thinking, after all, is common enough in life. I could even ask myself, hypothetically, what to do if in Caesar’s shoes at the Rubicon, knowing that in fact he went ahead and crossed. Hypothetical practice in deciding is one of the benefits of contemplating history—and the same goes for fiction. As for what I do when I plan for the case of being Caesar, we can describe it in various ways. I form a preference, we can say, between being Caesar and crossing and being Caesar and staying put. Or I make believe I’m Caesar and that it’s up to me whether to cross, rejecting some alternatives and permitting others. If there’s a problem with rejecting the alternative I know Caesar took, we can drop those features of the situation that might be up to me. The binge alcoholic can say on Wednesday, “If it’s forthwith to be Saturday night and it’s now up to me whether to take a drink, let me abstain.” He thus rejects drinking, in a hypothetical frame of mind.

We might worry that this changes the ordinary meaning of the term ‘plan’ so much that it’s now misleading for what I mean by the term. Perhaps so—but I really am appealing to the core of ordinary contingency planning: thinking what to do if a contingency arises, rejecting some alternatives, and allowing oneself others.

Once “plans” are understood in this way, can they act like beliefs? Might normative beliefs be plans or logically tied to plans, as I maintain? Reserve the term ‘ought’ as a quick way of saying “has most reason”. The unqualified dictum I started with was this: to believe that one ought to do a thing—that

² This follows the usage of decision theorists such as Savage, *Foundations* (1954).

one has most reason to do it—is to decide to do it. This Scanlon rejects, and rightly; it needs the qualifications I have just been stating, and which I stated in the book. My slogan to a closer and more verbose approximation might be this: to believe that a person ought to do a thing is to require it of oneself for the hypothetical case of forthwith being in that person's precise situation.

Deciding and planning, Scanlon says, are "things that one *does*, and that one can (in the full sense) only do for oneself." One can, though, as I have been saying, form preferences for the hypothetical case of forthwith being another; one can hypothetically reject or permit alternatives in that hypothetical frame of mind. Rejecting for oneself right now, moreover, can be treated as a degenerate case of rejecting hypothetically. It amounts, we can say, to hypothetically rejecting for the case of being oneself right now. If on Saturday night, the problem drinker rejects taking a drink for the case of being himself right then, this amounts to rejecting taking a drink right then. Rejecting for oneself right now is thus equivalent to a special case of something more general: rejecting hypothetically.

In short, then, the main thesis of the book is that we can understand normative judgments as an aspect of planning, so that every planner is committed to normative concepts and their logical features. I could make good the thesis, though, only with the right interpretation. If by a "planner" we meant someone who just picks a strategy, the thesis would be false, for reasons that Dreier, Bratman, and Scanlon all stress. Picking a strategy doesn't distinguish between thinking it strictly best and thinking it tied for best. If, though, we include in the notion of being a "planner" a difference between picking out of indifference and choosing out of preference, we have what we need, I claim.

Negation and Indifference: The Ideal Case

All the commentators worry that I might be taking by theft what I could acquire honestly only by cognitivist toil. This worry concerns the ideal case. What is the difference between indifference and indecision, Dreier asks, between allowing and mere failing to reject, between disagreeing and just not agreeing? First, am I not just helping myself to these distinctions without explaining them? Second, are they not best explained in cognitivist ways?

Dreier, like the other critics, gets the state of play in the book just right. Nicholas Unwin and others have argued that expressivists can't account for negation; we can't distinguish denying a claim and firmly suspending judgment.³ This problem ties in closely with indifference, with the problem of how to distinguish indifference from mere lack of preference. When I encountered these problems in the book, I was forced to speak of "disagreement", and to make such talk basic to the rest of what I was saying. I explained disagreement as best I could, describing the role it plays in a single person's

³ Unwin, "Norms and Negation" (2001).

thinking over time and in our thinking jointly when we “put our heads together”. I offered, though, no full explanation of disagreement—and it is hard even to know what a full explanation would consist in.

None of the critics question the intuitive notion, but Dreier and the others worry that the correct explanation of disagreement might not fit expressivism. “[I]t could be that the states with which it makes sense to disagree are the ones that are really representations of independent fact.”⁴ Likewise with indifference: to be indifferent between two alternatives, perhaps, can only be explained as thinking them equally good, where their being equally good is an independent fact.

I agree that to be indifferent among alternatives is to believe them equally good. Which way, though, can the explanation run? Could an independent fact, a fact of being equally good, explain the psychic state of indifference? Starting with such an independent fact encounters Moore’s notorious problem of saying what ‘good’ means, and so what it means to say that two alternatives are equally good. The same goes for reasons: If to be a “reason” is to count in favor, what is this “counting”? How can a consideration “count” in favor in a sense that is independent of whether anyone ever might count it in favor? We must either say something informative about what the special independent facts are, or turn non-naturalist, just claiming the facts to be *sui generis*. Naturalistic answers, we find, don’t work for meaning (though they must, I argued, work for such matters as what being equally good consists in). A heavy metaphysics of the non-natural layout of reasons—the “moron theory”, as Ronald Dworkin calls it—might propose a substantive explanation of the concepts if it worked, but the mystery of non-natural objects seems worse than the puzzle of the nature of indifference.⁵ Scanlon offers the approach of a “minimalist normative realist”, but if this consisted in offering no explanation of the concept, then it would just be theft in another form. (Scanlon himself is no such thief, and I address his conceptual holism later in these replies.)

I myself say that ideal planners serve to explain the features of these concepts. To establish this, I do need to appeal to an understanding of indifference or of disagreement. The cognitivist, though, has an alternative only when he can explain such things as the truth or falsehood of ought claims or reason claims. (And as for negation, orthodox semantics, remember, helps itself to the concept just as much as I help myself to disagreement. In this regard, the two sides engage in theft equally.) On my view, which Scanlon may share, talk of normative facts is explained by our concepts being in order. If the explanation is seriously meant to go in the other direction, normative facts confront heavy demands.

⁴ Dreier, this issue, p. 716.

⁵ Dworkin, “Objectivity and Truth” (1996), 104-105.

How far can disagreement extend? It does seem that I can come to disagree with a preference that I have had up to now—and to see this, we don't seem to need to settle whether preferences are in some robust sense judgments of special facts of preferability. We talk also of disagreeing with a decision; we can think a decision mistaken. Or a decision can be the only one open that wouldn't be a mistake. All these turns of phrase suggest that agreement and disagreement pertain not only to plain beliefs, but also to decisions and preferences.

Is disagreeing with a decision or preference sharply different from disagreeing with a belief? A decision, as Scanlon says, can't be true or false, and neither can a preference. That leaves us to ask, though, whether this is a symptom of deep import or just a matter of grammar. I suspect it is something of both. First, of course, sheer pickings can't be true or false, and that does tie in with the nature of picking. To pick an alternative, after all, isn't to disagree with picking another. Preferences too aren't correctly said to be true or false, but in their case the contrast may be more shallow. How significant is the distinction between thinking a preference mistaken and thinking it false?

A deep question does lie behind this one, I suggested in the book. Can we really disagree in preference? This, I said, amounts to whether to "put our heads together", whether to ponder jointly what to do if in a person's shoes, and why to do it. We can treat preferences either as no more than sheer personal characteristics, or as matters for dispute. If we treat them as subject to agreement and disagreement, I tried to show, we then mimic thinking in terms of oughts and reasons. Strands of ordinary thought, when systematized as Scanlon and I both think they should be in terms of "ought" and "reasons to do", can be explained as amounting to contributions to such joint thinking. (And the same goes for reasons to think or to feel.)

Once, though, we do help ourselves to disagreement and apply it to picking, we get all we need for the logic of normative concepts. To *reject* an act or a decision is to disagree with it. To *permit* the act is to disagree with rejecting it. When Buridan's ass is merely undecided, it follows, she doesn't disagree with eating either bale, and neither does she yet permit eating it. To permit eating it is to disagree with disagreeing with eating it. She moves to *indifference* when, for each of the two bales, she comes to permit eating it. Her preferences are *complete* when, for each alternative, she either disagrees with it or permits it.

As for "hyperplans", contingency plans (in my sense of the term 'plan') that are ultimately detailed, they aren't meant to alter the state of play concerning indifference or negation. The same goes for "hyperstates", combinations of a hyperplan with a complete view of how the world stands in matters of prosaic fact. These I meant as extreme idealizations of such ordinary combinations as a view on whether now to pack and a view on whether Moriarty

is nigh. Speaking in terms of them helps in establishing some of the conclusions I argue for. Dreier is quite right, though, that they don't change the status of the interrelated questions of what negation or disagreement consists in and how indifference contrasts with indecision.

A fully consistent planner in my sense of the term, I tried to show, would in effect deploy concepts that work much as a non-naturalist would think that normative concepts work. In particular, the concepts satisfy conditions of supervenience and factual constitution, in senses I explain in the book. I myself think that this reveals something basic to elucidating normative concepts. The commentaries in effect raise the question of what to make of this. From the ideal case we now draw the following lesson: A mere picker of strategies, as the critics stress, needn't have any ought-like concept. From what a planner picks, after all, we can't glean what he thinks he ought to do and what not. He might, after all, pick as he does but permit himself everything. Many different ought views, then, could eventuate in a given pattern of picking. A planner who can disagree with an act or a decision, though, has the logical materials he needs for something that acts much like our own normative concepts. These are conclusions we can draw when we consider only planners who are ideal.

Sartre and Incomparability

Are the concepts an ideal planner needs, then, our own, familiar normative concepts? Bratman and Scanlon focus on different grounds for answering no, Bratman on the "Sartre" case of existential choice, and Scanlon on acrasia. Bratman himself distinguishes three problems: Buridan, the Lady and the Tiger, and Sartre. The "Buridan" problem is the one I have been discussing, how to distinguish indifference from indecision. It arises for the case of the ideal planner. As for the Lady and the Tiger, I moved eventually, in the book, to subjective oughts, to the question of what one ought to do in light of one's information. That leaves the question of how to explain objective oughts, oughts in light of everything that is the case, whether or not the person has any way of knowing them.⁶

As for "Sartre", his character (Paul, I'll call him) found two kinds of considerations incomparable.⁷ How, Bratman asks, can this state of mind be distinguished from indifference or from indecision? Here again, we might try responding, disagreement is the key. Paul, imagine, chooses to tend his mother, but doesn't think that joining up with the resistance instead would

⁶ I discuss how to characterize objective oughts in terms of subjective oughts in "Truth and Correct Belief" (2005).

⁷ I use the term 'incomparable' for what philosophers often mean by 'incommensurable': neither better, worse, nor indifferent. The mathematical meaning of 'incommensurable' is different from this. I thank Mark Machina, who pointed out the *difference at a meeting* some years ago.

have been a mistake. If we can help ourselves to the notion of disagreeing, we get the following analysis. To think an act a mistake is to disagree with it. Does Paul believe it would be a mistake to think joining up mistaken? If not, it's a case of indecision: he neither disagrees with joining up nor permits it. If so, on the other hand, he permits joining up. Since he also permits tending his mother, we might try saying, he is indifferent between the two alternatives.

This last conclusion, however, is problematic. Incomparability is not indifference, for indifference is transitive whereas incomparability is not. Take two variants of tending his mother, one that includes eating breakfast today and the other not. Between either of these and joining up, Paul has no preference. Yet he prefers not skipping breakfast to skipping it. His lack of preference, then, can't be indifference.

This gives us a way, though, to distinguish incomparability from indifference, a way familiar in decision-theoretic lore. Not all patterns of permitting and rejecting generate complete preferences with transitive indifference, we are saying. Paul faces three alternatives: r , join the resistance, m , tend mother, and m^* , tend mother and skip breakfast. If all are feasible, he permits r and m , and rejects m^* . If r and m^* are feasible and m is not, he permits both r and m^* . No preference ordering is consistent with this, and we conclude that he regards r as incomparable with m and m^* .⁸

I'm not myself convinced, though, that such a state of mind is coherent. If not, it belongs at most in a discussion of planners who are non-ideal. What Sartre himself thought isn't entirely clear, for he said that in deciding, one "decides for all humanity."⁹ Perhaps to do this is to come to disagree with the alternative, to think it a mistake. Whatever Sartre thought, though, many will think that alternatives can be incomparable, so that a satisfactory account of normative concepts must allow incomparability as coherent to ascribe.

In *Wise Choices*, I spent considerable effort trying to distinguish the psychological states of thinking something to be a demand of rationality and taking it on as an existential commitment.¹⁰ There is much affinity between the things I try in *Wise Choices* and the proposals Bratman sketches in his commentary, and perhaps I should take some of his suggestions and return to something more in the spirit of my earlier treatment. As I say, though, I'm puzzled as to whether incomparability is a fully coherent notion. A central line of argument in decision theory shows that as an agent permits and rejects alternatives in a sufficiently rich array of contingencies, we can find utility and credence numbers such that it's as if those were her preference strengths

⁸ On deriving preference relations from choice functions, see Sen, *Collective Choice* (1970), Chap. 1*, sec. 6 (pp. 16-20). This violates Sen's condition β , p. 17.

⁹ Sartre, *L'Existentialisme* (1946).

¹⁰ In the year I first drafted large parts of *Wise Choices*, I had thrilling weekly lunches with Michael Bratman, and he pressed me repeatedly on these questions.

and degrees of belief.¹¹ As she permits and rejects alternatives, she comes to make the hard comparisons, and, we might say, ceases to regard alternatives as incomparable. Perhaps the right view to take is that as we stop being undecided, we come to make the comparisons, to believe, of two alternatives, that the first is better, worse, or equally good. That's the view I was drawn to in *Thinking How to Live*.

Paul acts, we are told, but doesn't think that the alternative would be a mistake. Perhaps he even forms preferences, but doesn't think that different preferences would have been mistaken. Is this coherent? He could think that which preferences to have depends on one's inclinations, or on one's past actions, decisions, or commitments. I'm not questioning the coherence of states of mind such as those. But can he think that even given his inclinations, commitments, and the like, neither preference would be a mistake? "Personalists" in decision theory do hold such a view, but there's a danger the view might lead to endorsing actions over time that are incoherent. Today, Paul boards the train with a plan to join the resistance. At his destination, his preferences change to another coherent form, and he boards the train back home. Once home, he reverts to his original preferences. At no point as he rides back and forth, we are supposing, does he make any mistake. (Eventually, to be sure, he loses grounds for any expectation that he will carry through on his plans—but could that be a hallmark of full rationality?) He spends his days and his savings riding the train, and neither joins the resistance nor tends his mother. Is this all coherent? (Buridan's ass, once she takes a step toward a bale of hay, will find going for that bale more desirable, since it now takes fewer steps to get to it. One step tips the balance. Incomparability isn't the same as exactly even balance, and a step or even a train ride may not bring with it a decided preference for the alternative approached, even though it is now less expensive than it was.)

If we do have to accommodate incomparability as coherent, then I ought to welcome Bratman's rich array of proposals and return to something more in the spirit of my treatment in *Wise Choices*. Disagreement remains part of the key, but it might come in various grades, for purposes of current planning and for full interpersonal discussion. I hope to avoid all this, but the phenomena might demand it.

The highest grade of disagreement might then, as Bratman suggests, require something like a plan to defend in joint inquiry into how to live. This, however, he notes, would rule out a judgment individualist. I don't find this immensely worrying. The judgment individualist, it seems to me, is someone who conducts conversations with himself in a spirit he learned in interpersonal conversation, but disciplines himself not to slip into doing that

¹¹ Hammond, "Consequentialist Foundations" (1988) formulates such an argument in terms of contingency plans and their properties.

with other people. No straightforward explanation of this state of mind can be given, but perhaps it isn't a straightforward state of mind.

Weakness of Will

A core part of the book was a possibility proof, showing that an ought-like concept does emerge from planning. That left the question of whether this is a concept we have. There may not be any clear answer, since ordinary terms like 'ought' come with many shades of meaning, mostly not all that sharp. I adopted Ewing's proposed sharpening, where a primitive ought is the basic conceptual atom that gives normative concepts their special character. A reason to do something (or a "reason to", as I'll put it) is then a consideration that counts toward an ought. On these systematizations of ordinary thought Scanlon and I agree. We disagree, however, on whether these concepts can be explained as emerging from planning in the way I describe. Does thinking that one "ought" to do a thing, in Ewing's primitive sense, consist in requiring it, in the sense of 'require' I have developed? Is the strength of a reason a decision weight, so that one regards something as a reason to the degree that one weighs it toward doing the thing?

Scanlon, as I understand him, thinks that even if the possibility proof works (which he thinks it doesn't), it can't explain the ordinary concept of a reason. The counterexamples stem from *acrasia*. If this is right and if the possibility proof goes through, then the concepts that would fill the needs of an ideal planner aren't concepts that we imperfect planners use.

If the problem drinker truly rejects taking a drink and keeps on rejecting it, he doesn't drink. A person can, though, Scanlon would insist, believe that he ought not to drink and still drink. He can do so even for the special sense of the term 'ought' that I'm now using: he can think that he, at this very moment, has most reason not to drink, and still take the drink. To believe that one has most reason not to do a thing, then, can't be to reject doing it—hypothetically or for one's actual case. *Acrasia*, it seems, shows any analysis in terms of planning to be false.

If a person fails to do something, I responded in the book, then at least he isn't "of one mind" in believing that he ought to do it. Scanlon denies this too, citing the example of failing to call the doctor about a disturbing symptom. I myself don't find this case clear, and I wonder whether a clear case can be found for Scanlon's contention. My belief, to be a counterexample, must be that I have most reason to call the doctor this very moment. Now I can easily picture thinking that I ought to make the call very soon, but at this particular moment I'd better wait. (Kant speaks of making an exception in my own case, and I might likewise make an exception for this very moment.) I can all too easily picture doing this repeatedly during the day whenever I bring myself just to the point of making the call. At the moment I'm walk-

ing away from the phone to the refrigerator, though, should we really say I'm of one mind in thinking I ought not to be doing what I'm doing? Even if I yell to myself as I leave the phone, "At this very moment, I have most reason to call!" am I of one mind in believing what I yell? I'm not sure.

In *Wise Choices, Apt Feelings*, I fixed on acrasia and tried to accommodate it. I speculated on a special psychic state that I called "accepting" a norm, picturing it as the work of a specially human, language-infused motivational system. This system could be overpowered by other motivational systems, some evolutionarily older and others intensely social. That may still be the best way to treat the matter, but it raises the worry that our motivations may not turn out to divide so neatly. In *Thinking How to Live*, my approach was different. Ideal thinker-planners, I tried to show, would have a concept that looks like Ewing's primitive ought. The tie to imperfect reasoners like us was then bound to be messy, but the problem, I thought, is the general one of how conceptual features tie in with human performance. Cases described as acrasia, I came to think, are cases of vacillation, unclarity, or inconsistency. Fitting such a case into the rubric of acrasia, as a case where some clear belief with clear content is at odds with one's actions, isn't true to the phenomena.

If I now grant Scanlon that the rubric does sometimes apply, how much of what I thought must I give up? My hypothesis was that, as a matter of their logic, the two sets of questions are the same, ought questions and questions of what to permit and what to reject. There isn't a realm of theoretical reasoning about reasons and their strengths and the resultant oughts, distinct from practical reasoning about how to weigh the considerations and what, as a consequence, to do. This hypothesis leaves open the question of how we are to be interpreted as wielding the concepts we share with an ideal planner. We think and sometimes speak our thoughts, and we act. In interpreting us, much weight goes to the pattern of our words. In our actual, confused motivations, considerations don't have weights that compose precisely. We may, though, have words that approximate the words an ideal planner would use, and that tie to action roughly as they would in an ideal planner. Those words we can interpret as ones the ideal planner might use. The ideal planner always does what she thinks she ought to do. We ourselves may not have words, in our thought and talk, that both fit the pattern of the ideal planner's words and tie invariably to motivation. And so we may have a belief that is best interpreted as an ought belief, but not do the thing we are best interpreted as believing we ought to do. I may be best interpreted as believing that I ought to call the doctor this very moment, even though I don't start calling.

This is close to what Scanlon maintains. If I allowed acrasia in this way and Scanlon accepted my possibility proof, we would then agree that an ideal planner reasons correctly with ought, and does what he concludes he ought to do. We would agree that his reasoning with ought would exactly match his

practical reasoning, in the sense of reasoning what to permit himself and what to reject. We would agree that our own ought reasoning ought to match correct practical reasoning.

Still, there might remain differences. My ideal planner is in some ways a simple being, perhaps even a crude automaton, who lacks the mental richness needed to be attributed a human-like grasp of human concepts. For the ideal planner to have the full concepts of ought and reason, doesn't she need far more than I'm allowing? In a way, I agree that she does. The computer on my desk doesn't really grasp the concept *and*, in that it can't think: nothing in it has the right kind of ties to a human-like richness of experience and thought. Still, though, the logic of the concept *and* is given by simple rules that can be represented on the stupidest computer.

What features of our use of terms like 'ought' and 'reason' form part of their logic? One candidate would be substantive features. A person doesn't have the concept of a reason, we might insist, if he can't be interpreted as thinking that the fact that an act would lead to suffering is a reason not to do it. Or at least he must be interpretable as accepting the preponderance of a large number of such truisms. Now if the question is simply how to interpret our ordinary use of 'ought' and related terms, then there might be much to be said for this—and it might support Blackburn's contention that a person can count as meaning "ought" in the ordinary sense even when he disagrees with the rest of us both on extension and in attitude. For philosophical purposes, though, it may be best not to place such a requirement on the concept of a reason. Otherwise, we get perverse answers to some of our why questions. "Why avoid suffering?" you ask. "Because suffering provides reasons"—pretty empty, but so far so good. "You wouldn't count as wielding the concept *reason* if you weren't best interpreted as thinking that suffering provides reasons." Now things have gone off track; this isn't relevant.

Scanlon mentions another feature of the richness of human thought about reasons: its role in criticism. Part of its role in criticism, I answer, stems from its role in planning, and the other parts aren't a matter of the logic alone. A fully human use even of the concept *and*, after all, includes its role in criticism. That doesn't make the logic textbooks incomplete in their treatment of the concept.

The tie of ought beliefs to action is normative, Scanlon proposes. A person who fails to do what he thinks he ought to do isn't fully rational. I agree with this dictum, of course, but all that's substantial in it we can glean from the simplistic ideal planner of my story. On some readings, this dictum would be trivial. We could stipulate that acting on what one takes to be the preponderance of reasons is part of what the term 'rational' means as applied to people. The dictum is then empty: The person who is fully rational, we are saying—meaning the person who, among other things, always does what

he concludes he has most reason to do—always does what he has most reason to do. Perhaps instead the dictum gets its content from the rich array of attitudes that tie to shortfalls of rationality in people. Finding myself short on rationality, I can kick myself in disgust or form a resigned smile. Catching someone else, I can feel triumphant and cry “gotcha!” or I can try to be helpful. All of these reactions, though, come on some occasions and not others. None, it seems to me, qualify as built into the meaning of ‘rational’.

What holds, then, as an invariant matter of conceptual logic? If a person joins believing he ought to do a thing with failing to do it, we can say, he does something he ought not to do. For this he can be criticized invariably. Whatever else the criticism involves—humiliating him and calling him a fool and an idiot, or sympathizing with the human condition—the criticism implies that he ought not to join the two. He ought not to join believing he ought to do a thing with failing to do it.

Why is it incoherent to deny this dictum? For a reason that any reasonable account easily predicts. Hamlet thinks he ought call the doctor right now, but doesn’t call. It follows that either (i) he’s mistaken about what he ought to do, or (ii) he’s doing something he ought not to do. Either way, he’s subject to criticism. Either way, he’s believing or doing something that he ought not. One can’t coherently agree with both the belief and the action. This, then, is the intelligible and substantive normative tie of ought belief to action. (And the tie of plan to action is, likewise normative, we might try saying: if Hamlet plans right now to call the doctor but walks away from the phone, no one could agree with both his plan and his action. We disagree with the package even before we settle what to do if in his shoes—and that’s to criticize.)

The akratic agent, on the account I’m now trying out, is inconsistent. He reasons to a conclusion as to what to reject, and yet doesn’t really reject it. He’s like a person who reasons to the conclusion that he is mortal, and still thinks that nothing can kill him. A normative tie to action, then, is built into the concepts of ought and reason. The norms are the ones built into the concepts.

This tells us what we are asking when we ask about oughts and reasons. Ought questions are practical; they are questions of what to require of ourselves. As for reasons, to ask how strong a reason is, we can say, is to ask what weight to give it in one’s decisions. We can’t find two clear, distinct questions to ask ourselves, first, how strong a reason a consideration is, and second, what weight to give it in decisions. The ideal thinker-planner need make no such distinction, and it would be bizarre if we need a concept that she would find superfluous just to give us another way to be irrational. Still, we can perhaps sometimes best interpret ourselves as thinking, in a confused

state of thought and action, that we ought to be doing a thing that we aren't doing. Whether this last is so I'm not sure.

Properties and the Aim of Reason

How distinct is my own view from a realism like Scanlon's? I began with a slogan: Thinking what I ought to do is thinking what to do. Blackburn and Sinclair point out, though, that a realist too could accept this slogan. Indeed I can think of more than one line that such a realist might take. One way would be to argue that "what to do" just means what one ought to do.¹² To be a realist on what one ought to do is thus to be a realist on what to do, the critic can say, and the ought property is the starting point of explanation. I myself, though, intended "what to do" to tie to action, so that concluding what to do entails deciding, at least hypothetically, and so that to conclude what to do right now, at this very moment, involves settling on doing it. The starting point for explanation is deciding—or at least, as I keep saying, deciding up to the point of indifference. Even if I concede to Scanlon that we might sometimes be best interpreted as employing the ideal planner's concept even when the tie to action fails, the point remains that the starting point for explanation is agreeing and disagreeing with actions.

A second way to be a realist and accept the slogan might be the one that Blackburn and Sinclair themselves broach: to claim that there is "a single property that makes it correct to have a certain intention." In a sense, though, I myself accept this last claim. The property in question, I say, is that natural property, whatever it is, that constitutes being okay to do. If, for instance, hedonistic egoists are right, it is the property of maximizing one's prospects for net pleasure. Unlike an analytical naturalist, though, I deny that when we ask which property this is, the question is semantic or conceptual. It's a planning question, a question of how to live. One comes to a view on the matter in the course of deciding aspects of how to live.

Is the link of the concept of ought to this property, then, "a metaphysical link, not itself the creature of expressed attitudes," as Blackburn and Sinclair tell us a realist might maintain? I don't know what a metaphysical link would mean, and the story I tell explains the link with no talk of metaphysics. Is the link of concept to property, then, instead "the creature of expressed attitudes"? If no one had ever expressed attitudes or even had them, the link would still be there—or so it's consistent to claim.¹³ For all the sheer meaning and logic of 'better' tells us, lush forests might be better than barren wastelands even if no one were there to appreciate them.

Do ought judgments, on my view, *track* a property? Not in the sense that the concept *ought* plus the layout of properties in the world suffice, together,

¹² This was suggested to me by Jason Stanley.

¹³ See Blackburn, "How to be" (1988), p. 172.

to settle which property constitutes being what one ought to do. Which property it is, as I say, hinges on how to live—a question one comes to views on in the course of making decisions. On the other hand, I do think there is a property that, by *a priori* necessity, an act has just in case one ought to do it. If that counts as tracking, I do think that ought judgments track a property.

These shades of interpretation aside, Blackburn and Sinclair are on target on the main contention that distinguishes my own position from a realist's. The property that constitutes being what one ought to do, I hold, doesn't itself figure in the explanation of the concept *ought*. There is such a property: anyone who plans and decides, I argue, is committed to this claim. The property, though, is not where to look in order to explain the concept. Look instead to planning and decision.

What distinguishes the moral use of 'ought' in particular? As Blackburn and Sinclair complain, I scarcely discuss the question in this book. Rather, I rely on what I have said before on the topic.¹⁴ Morality, as opposed to normativity in general, might be distinguished in various ways, but I have argued for the special importance of a narrow sense that is tied to the moral sentiments—to feelings of guilt and resentment. This explains how morals bear on what to do. The bearing is indirect: moral judgments concern directly how to feel about actions, and the feelings tell for or against. I still think that this is a helpful way of thinking about morality, but it needs refinement. Howard Nye has stressed to me that guilt is retrospective, whereas morality must be tied to aversions to things one hasn't yet done but might do. I need to speak of a special kind of guilt-tinged aversion involved in narrowly moral motivation—along with an attraction to doing the right thing, as one sees it, that is tinged with self-approbation. I have also, since this book, had second thoughts on the psychological reality of my earlier treatment of narrowly moral concepts—though I still think that the most defensible concepts work in the way I describe, so that those are the concepts to credit to ourselves out of charity.¹⁵

One more matter of interpretation: what is the goal of practical reasoning? It's not quite what Blackburn and Sinclair say I think it is, "to form a fully worked out contingency plan." The goal, I can only say, is to figure out what to do. You achieve this goal if you settle on doing something that really is the thing to do, or at least is okay to do. You could form a fully worked out contingency plan, though, and fail in this goal—if, for instance, the plan said to maximize your own expected prestige, whatever the costs to

¹⁴ Moral concepts are a chief topic in *Wise Choices, Apt Feelings*, and I discuss the issue further in "Moral Concepts" (1992).

¹⁵ The second thoughts and rethinking of the status of my claims is in "Moral Feelings" (2006).

your own happiness and that of others. Not that you would then fail in your own eyes: if that were your plan, you yourself would think that a person ought above all to garner prestige, and so you would think that you had successfully figured out what you ought to do. But you would be wrong. In saying this last, note, I'm departing from strict metatheory and stating a substantive normative view, a view on how to live. The metatheoretical point to make is that this view is coherent, and so are alternatives. Whether you would be right or wrong, then, isn't settled by the nature of the concept *ought* alone. Would you, in this case, have attained the goal of practical reasoning? That is a planning question, a question of how to live.

References

- Blackburn, Simon (1988). "How to Be an Ethical Anti-Realist". *Midwest Studies in Philosophy* **12**. Cited as in Blackburn (1993), *Essays in Quasi-Realism* (New York: Oxford University Press), 166-181.
- Dreier, James (1996). "Expressivist Embeddings and Minimalist Truth". *Philosophical Studies* **83**, 29-51.
- Dworkin, Ronald (1996). "Objectivity and Truth: You'd Better Believe It". *Philosophy and Public Affairs* **25**, 87-139.
- Gibbard, Allan (2005). "Truth and Correct Belief". *Philosophical Issues* **15**: *Normativity*, 338-350.
- Gibbard, Allan (1992). "Moral Concepts: Substance and Sentiment". *Philosophical Perspectives* **6**, James E. Tomberlin, ed. (Atascadero, CA: Ridgeview Publishing Co.), 199-221.
- Gibbard, Allan (2006). "Moral Feelings and Moral Concepts". In *Oxford Studies in Metaethics I* (Oxford: Oxford University Press), 195-215.
- Hammond, Peter (1988). "Consequentialist Foundations for Expected Utility". *Theory and Decision* **25**, 25-78.
- Ramsey, Frank Plumpton (1931). "Truth and Probability". *The Foundations of Mathematics and other Logical Essays* (London: Routledge & Kegan Paul).
- Sartre, Jean-Paul (1946). *L'Existentialisme est un Humanisme* (Paris: Nagel).
- Savage, Leonard J. (1954). *The Foundations of Statistics* (New York: Wiley).
- Sen, Amartya K. (1970). *Collective Choice and Social Welfare* (San Francisco: Holden Day).
- Unwin, Nicholas (2001). "Norms and Negation: A Problem for Gibbard's Logic". *Philosophical Quarterly* **51**:202 (January), 60-75.