On Thinking How to Live: A Cognitivist View

Philip Pettit

Thinking How to Live, by Allan Gibbard. London: Harvard University Press, 2003. Pp. xv + 302.

Allan Gibbard's strategy in his new book is to begin by describing a psychology of thinking and planning that certain agents might instantiate, then to argue that this psychology involves an 'expressivism' about thought that bears on what to do, and, finally, to try to show that ascribing that same psychology to human beings would explain the way we deploy various concepts in practical and normative deliberation. The idea is to construct an imaginary normative psychology, purportedly conforming to expressivist specifications, and then to campaign for the hypothesis that that psychology is ours and that we ourselves conform to those specifications.

The upshot is an original and intriguing argument for the claim that 'expressivism', as Gibbard understands it, is sound. There is more in the book than that bare argument. Filling out the strategy pursued, for example, the volume contains useful discussions of a number of current debates (chs 2, 12), an enlightening argument that the strategy survives the introduction of the notion that a concept can have a character distinct from its content (ch. 6), and an extended set of reflections on how far the point of view defended leaves room for the idea of gaining normative knowledge (pt. 4). But I shall ignore those aspects of the book in this commentary; the focus is on its central argument.

My approach will be to reconstuct the main stages in the *Aufbau* of our normative psychology that Gibbard provides, emphasizing some crucial points that are passed over rather quickly, and then to suggest on the basis of this reconstruction that the more natural lesson to derive is not an expressivist one. Although I break with Gibbard in that way, however, I have great admiration for the book. *Thinking How to Live* is a sharp, fresh and invigorating treatment of the main issues of metaethics.

Mind, Vol. 115 . 460 . October 2006
doi:10.1093/mind/fzl1083

My discussion falls into ten sections. I develop my account of Gibbard's *Aufbau* in the first eight sections. On the basis of features emphasized in that account, I argue in the ninth section that the story provides support for a cognitivist rather than an expressivist point of view. And then in the tenth section I reinforce that argument by showing that Gibbard's position remains vulnerable to a problem that Frank Jackson and I posed some years ago for expressivism.

1. Representations and plans

Let us begin, Gibbard suggests, with a picture of some thinkers and planners who, like us, reason their way to decisions (p. 5). These agents will have to be equipped with representations of facts about their world, about themselves, and about what they are in a position to do in changing the world. These representations, which Gibbard describes as prosaic beliefs—I shall sometimes speak of beliefs, period—are beliefs in indisputably factual contents. They are needed to provide the agents with guidance on the obstacles and opportunities they confront as they seek to pursue one or another goal. They can do that so far as they are suitably tailored to the facts they register, coming and going as the evidence stands in their favour or stands against.

But apart from representations, of course, any agents will also need to have goals and it is here that Gibbard's scenario becomes distinctive. He stipulates that the agents he imagines are to be equipped with planning as well as representational states of mind. It is these planning states that provide his agents with their goals.

For Gibbard, a decision as to what to do now—the formation of a state that selects one alternative among those available by the agent's lights—involves planning. But so, secondly, does the formation of a state that would select one of the available alternatives—or that would at least identify a subset of selectable alternatives—in a hypothetical situation, expected or otherwise. And so, thirdly, does the formation of a state that would serve this function for another agent in an actual or hypothetical situation. One may plan for now, for later, or for purely imaginary situations, and one may plan for others as well as for oneself. As 'representation' may refer to a state of representing or to a content represented, so 'plan' may refer to a state of planning or to the content of what is planned: the goal that is pursued (p. 47). Taken as a state of planning, a plan amounts to just 'a determination of what to do in various contingencies, expected or hypothetical' (p. 53), where the agent targeted may be oneself or another.

In postulating representations and plans, we have not said anything on how complete they are. I may represent various ways the world is but I certainly do not represent all. And equally I may plan for various situations involving myself or others but I certainly do not plan for all. The schema of representation and plan allows, however, for extension to the limit case of an agent who represents all the ways that the world is and might be, and who decides or plans for all the situations possible and all the agents possible, under that way of viewing the world. To instantiate such an extended mentality—in particular, one that will serve the agent unproblematically in every possible choice—is, in Gibbard's terminology, to embody a hyperstate of representation and planning. The limit is impossible, of course, but it gives Gibbard a way of depicting the ordinary thinker-planner as an adumbration of certain hyperdecided counterparts. The counterparts of an agent will be those more than human subjects—Gibbard often personalizes them as Zeus and Hera—in hyperstates that the agent could reach without changing his or her mind (p. 91).

2. Rationality

Two representational states may combine with one another unproblematically, making it possible to be guided at once by both, or they may fail to combine in that way: they may rule one another out in the manner of believing that p and that not p. Similar combinatorial restrictions, Gibbard now points out, also govern planning states. It will be possible to combine some plans with one another unproblematically, as in planning to go to town and then to do the shopping. It will be impossible to combine others in that way, as in planning to go to town and planning not to make any arrangements for getting there. Planning to go to town does not rule out doing the shopping but it does rule out neglecting to make arrangements to get there; and as it rules out neglecting to make any such arrangements, it rules in making some.

These examples all involve plans for oneself but parallel constraints will apply across plans that involve others as well as oneself. It will make no sense to plan on doing something oneself, for example, that does not fit with what one plans for some others to do, although one may also have back-up plans for what to do in the event that others fail to act as one plans for them. This parallels the way in which one may make plans for what to do now that fit with what one now plans to do tomorrow, while having a back-up plan for what to do the day after

tomorrow in the event of not having acted as one now plans to act tomorrow. There is a question, of course, as to how one should plan if one believes that in some domains other agents, or even one's later self, will generally not act as one plans for them to act. But Gibbard does not discuss this and I shall not dwell on the issue here.

Not only will representations connect under certain restrictions with representations, and plans with plans; they will also connect in a corresponding way with one another. My representation that it is already dark will not combine unproblematically, for example, with my plan to go to town before nightfall. Thus when Gibbard asserts that 'certain states of mind have inferential import in that they rule out other states of mind' (p. 45), he means states of mind to cover either representational or planning states.

The inferential import of such states will be sourced in the consistency and inconsistency, the compossibility or non-compossibility, of corresponding scenarios: that is, in presumptively objective and inescapable conditions. If it cannot be the case both that p and that q—if these scenarios are not compossible—then I cannot unproblematically form the representation that p and the representation that q and I cannot unproblematically plan on making it the case that p and on making it the case that q. The one representation or plan will rule out the other, and forming one or other will rule in the eschewal of the alternative. This comment on compossibility is mine but I assume that Gibbard would endorse it.

The combinatorial restrictions on representational states means, in familiar terminology, that there are constraints and ideals of coherence or rationality available to govern those states; strictly, constraints and ideals of deductive as distinct from inductive or evidential rationality. These are standards of functioning such that any shortfall means that the representations will not fully or properly do the job of guiding the agent. Rationality will be flouted so far as an agent instantiates representations that rule one another out, for example; and rationality will fall short of full satisfaction so far as the agent does not instantiate a representation that existing representations rule in (or does not reject that representation and revise existing representations). What is true of representations on their own is also true of plans on their own and of representations and plans together. In each such case the existence of the combinatorial restrictions means that rational constraints and ideals are available to gauge an agent's performance; these will be standards of deductive rationality, broadly conceived. Let a putative agent prove too irrational, of course, and we will have doubts as to whether

there is a real agent present, but we normally do not shrink from allowing agents to breach intuitive constraints of rationality in some degree and still count as agents (p. 12).

In order to articulate the demands of rationality on a given thinker-planner, Gibbard invites us to consider the hyperdecided counterparts of the agent. These agents will be rational so far as their extended mentality, as I put it earlier, will serve them unproblematically in every possible choice. On this assumption, Gibbard points out, a thinker-planner's set of representations and plans will be consistent, if there is at least one counterpart, and so one hyperstate, that answers to the given set; it will be inconsistent if there is none (p. 59). And a set of representations and plans will commit an agent rationally to endorsing a further representation or plan, so he claims, if this is endorsed by every one of the agent's counterparts (p. 91).

3. Agreeing and disagreeing

One of the important aspects of the psychology described by Gibbard, given his purposes, is that it enables us to see that different agents can agree or disagree, not only in their representations of how things are, but also in the plans they form. This is important because it means, to anticipate a little, that when thinker-planners express rival plans then, as we would expect, they can be genuinely at odds with one another.

You and I will disagree in representation just in case I represent the way things are in one way, you in an incompatible manner. But you and I can also disagree without any disagreement about how to represent the world. We will disagree to the extent that what you plan for yourself to do in some contingency does not coincide with what I plan for you to do: what I plan for you as my targeted agent. The possibilities on this front are limitless, according to Gibbard, since I may disagree or agree with anyone about any particular decision. It may even be, for example, that 'I disagree with Caesar on whether, if he in his shoes, to go to the Senate on the Ides of March' (p. 68).

The sort of disagreement that Gibbard has in mind is not disagreement about whether going to the Senate was the option that fitted with Caesar's background plans. I may believe—believe in the ordinary prosaic sense—that Caesar's background plans did not rule out that option, or even that they ruled it in. Thus I may think that relative to his plans—taking those plans as given or unquestioned—the decision to go to the Senate was rational: rationally permitted or even required. But, notwithstanding this convergence in our representations of what

was rational for Caesar, I may still disagree with him in Gibbard's sense: I may plan if in his place to stay at home. Caesar's plan to go to the Senate and my corresponding plan differ so far as his bears on what he should do and mine bears conditionally on 'what to do *if* in Caesar's shoes' (p. 69).

As I may disagree or agree with another person in this way, of course, so I may disagree or agree with myself across time. I may agree with an earlier self in representing what I did as rational in the light of my existing states at an earlier time. But I may still disagree with what I did at that time. I may have a different conditional plan now as to what to do if in my shoes then from the plan that I then had as to what to do in my shoes at that time.

Why should we be interested in interpersonal or indeed intertemporal agreement and disagreement? To the extent that we can communicate such agreement and disagreement—more on this later—it may be to our evolutionary advantage; and in any case, it is likely to be personally helpful. 'In conversation, I may let you do some of my thinking for me. I stand ready, of course, to disagree with you if I find disagreement called for—just as in my own thinking I stand ready to disagree with a thought of five seconds ago' (p. 71). Deliberating our rival plans for what I do or for what you do 'can make us better deciders than we would otherwise be' (p. 53). To invoke a notion to which we turn later, it can enable us to reason with one another, generating the traditional benefits of joint deliberation.

4. A limitation of the psychology

The presentation I have just given of Gibbard's initial scenario departs from his in making use of words like 'representation' and 'rationality' and in introducing talk of compossible and non-compossible scenarios. But the presentation is still entirely faithful, I believe, to his picture of the thinker-planners with whom he wants to begin. The reason I adopt it, rather than sticking more closely to his chosen words and chosen emphasis, is that it brings out the following, easily missed feature. Under the picture presented, the plans will have to materialize and evolve in interaction with one another and with representations, but for all that has been said so far they may do so without the agent being able to form any representation of the connections involved or of the relations of compossibility at their origin.

The feature can be highlighted by a comparison with decision theory. Under Bayesian decision theory, in any of its many versions, agents are

depicted as having credence functions and utility functions of such a kind that any alternative preferred among a given set of options will maximize the agent's expected utility. That alternative will have outcomes, according to the agent's credence function, such that if we add the utility of each outcome times its probability to the utility times the probability of every other outcome, then the sum is greater than the corresponding sum for any other alternative; I abstract away from the different ways in which the elements are construed under different versions of the theory. The main theorem of decision theory is that any agent will maximize expected utility in decision-making if his or her preferences over alternatives, including preferences over complex alternatives like gambles, satisfy certain conditions.

A striking feature of decision theory is that when an agent is moved to pick an option that maximizes expected utility, the agent will not be required to form a credence in the fact that this option has that attraction. Suppose we construe the credences of decision theory as degrees of belief and the utilities as degrees of desire. Under that construal, the theory certainly suggests that when rational agents choose an option that maximizes desire-satisfaction according to their beliefs, they do so because it has that attractive property. But, crucially, it does not suggest that they do so because of believing at a higher level that it has that property. The decision in favour of the option materializes because of its connection with the agent's beliefs and desires but the connection need not be something about which the agent forms a higher-order, reflective belief. The decision materializes rationally, maximizing expected desire-satisfaction, but it need not materialize in a way that requires the agent to believe that it has that feature (Pettit 2002, pt. 2).

Gibbard's plans are unusual in extending from the actual to the possible, and from oneself to others. But for all that has been said on this point in introduction of the theory, it may operate in a parallel manner to decision theory. The plans may materialize and evolve under the rules governing what can be combined with what—under the rules deriving from facts about compossibility—and so under the constraints and ideals of broadly deductive coherence or rationality; they may combine appropriately with one another and with various representations. But they can materialize and evolve rationally in this manner without planning agents forming or being guided by representations to the effect that this or that scenario is compossible or not compossible with some other, or that this or that plan is rationally permitted or required. It is one thing to be constituted so that plans will unfold and develop in a way that is rational overall. It is quite another

thing to be constituted so that this happens under the surveillance of representations to do with compossibility or rationality. Gibbard's story up to this point requires the achievement of a certain degree of rationality by a thinker-planner but it does not require any awareness or representation on the agent's part of what this achievement involves.

Another way of putting the point is this. In order to be deductively rational, a thinking-planning agent will have to respect the facts of compossibility, and conform to the rules of rationality. But an agent may respect such facts and conform to such rules without having the concept of compossibility, and so without being able to represent any option as permitted or required. The agent may conform to the rules of rationality in forming this or that plan or representation without the conformity being in any way explained by the predication of such a higher-order property. The property of compossibility is meta-propositional in the sense that it applies to the potential contents of representations and plans—that is, to propositions. Lacking the concept for such a meta-propositional property, an agent may conform only blindly to rules of rationality. And if blindly, brutely: that is, without a conception of the requirement playing any explanatory, regulative role.

Gibbard thinks that his thinker-planners are agents and like us, therefore, in so far as we are 'beings who can reason our ways to decisions' (p. 5). But for all that we have seen up to this point, his thinker-planners may only be able to reason in a Pickwickian, purely functional sense. While they instantiate a process whereby rationality is reliably achieved, the process may be entirely blind and brute: it may be beyond their conceptual reach and outside their regulative control. Reasoning in the ordinary sense presupposes the capacity to apply meta-propositional concepts like that of compossibility and to find a use for expressions like 'so' and 'therefore', and cognate terms. Gibbard's thinker-planners, for all we have seen so far, may be incapable of reasoning in that sense about what to do or even what to think. They may be instances of *homo rationalis* but they need not be examples of *homo ratiocinans*.

The limitation described here is important but should not be overstated. Were thinker-planners to develop meta-propositional concepts, they could seek out information on what is compossible with what, and so on what it is rational to hold or plan, thereby generating extra checks on what they came to believe and intend: that is, they could reason in the familiar sense of the term. More on this later. But of course the reasoning could not go the whole way up in the hierarchy of their meta-propositional beliefs, as Lewis Carroll (1895) taught us. The belief that it

is rational to believe something, for example, must be capable of generating that belief all on its own, in virtue of the make-up of the agent. On pain of an infinite regress, the generation cannot rely on the person's having a yet higher-order belief to the effect that it is rational to form a belief that one believes it is rational to form; or if it does, it cannot rely on the presence of a yet higher-order belief; and so on: at some stage the regress must stop.

5. The turn to expression

At this point in his construction of thinker-planners, Gibbard makes a crucial move. He postulates that his imaginary subjects develop a term and a concept for that alternative or subset of alternatives in any situation that their planning states leads them—presumptively in a rational way—to select, whether for themselves or others. He imagines them using the expression, 'the thing to do', to identify a planned-upon alternative: one, presumptively, that is uniquely permitted under their planning states, and 'okay to do' to identify an alternative not ruled out by their plans: one, presumptively, that is permitted under their planning states (pp. 7, 88). Gibbard insists that in envisaging this development, he is not breaking with expressivism. 'I appropriate the phrase "the thing to do" to serve as an expressive term—ignoring all question of what this phrase means in normal English' (p. 7). And the same, he suggests, is true of 'okay to do'. Whatever its ordinary connotations, we are to take each phrase in a technical, expressivist sense.

But what sense exactly is that? In illustrating the idea, Gibbard says: 'to conclude, say, that fleeing the building is *the thing to do* just *is* to conclude what to do, to settle on fleeing the building'. While we may describe the decision involved as a case of coming to think that fleeing is the thing to do, the idea is that the thinking and deciding are not 'separate activities'. In particular, the thinking does not count as an instance of representing, only as a case of deciding or planning (p. 13). 'By sheer stipulation, then, the meaning of this phrase "the thing to do" is explained expressivistically: if I assert "Fleeing is the thing to do", I thereby express a state of mind, deciding to flee' (p. 8). I express that practical attitude and, crucially, I express nothing more: not a self-focused belief that I have that attitude, and not a world-focused belief that fleeing has a subjective profile—a to-be-doneness—that the presence of such an attitude might be thought to give it.

What does it mean to express such a practical attitude? 'To express a state of mind', Gibbard says, 'is to purport to have it, whether or not

one does' (p. 77); it is to try to get others to believe that one has the attitude and, presumptively, to invite them thereby to rely on one's having it. In that sense someone who says that fleeing is the thing to do will certainly be purporting to have the corresponding plan for the targeted agent. In using 'the thing to do' of fleeing, Gibbard thinks that one will be prompted in the first place by the content of the plan—the *agendum* envisaged—rather than by an awareness of the existence in oneself of the corresponding state. In that sense, as he emphasizes, his expressivism is projectivist, like Simon Blackburn's (1984). Let plans materialize so that a given alternative stands out as uniquely eligible and is selected by the agent. The projectivist idea is that the phrase 'the thing to do' will go with the way that that alternative then presents itself, as a regular phrase like 'the biggest thing around' will go with the prosaic, descriptive way that a regular object may present itself.

It may seem, of course, that in order for me to pick out an alternative as 'the thing to do', holding or making a plan to do it, I must form a representation of it as the thing to do—a representation expressed by saying it is the thing to do. But Gibbard insists that while there will be a belief present to guide the use of a descriptive expression like 'the biggest thing around', there need be no belief or representation present in his thinker-planners to guide the use of a phrase like 'the thing to do' or 'okay to do'. Finding an alternative to be the thing to do, or okay to do, is not a matter of ascribing a property to the option in the way that a belief or representation would ascribe a property, even though those terms and concepts answer to the way options are seen by the agent. There are no such properties in the world that are available to be ascribed: no intrinsic to-be-doneness or okay-to-doness.

If thinking that something is the thing to do just is a decision or plan to do it, of course, then thinker-planners cannot ever fail to do what they take to be the thing for them to do. And in that case are they not going to be very different from the weak-willed human beings whom they are supposed to model? Gibbard's answer is that the thinker-planners, like us, may often be of many minds, with one mind planning to do something, and expressing this in a sincere comment on the thing to do, while another mind takes a different attitude and may win out in the end (cf. Stalnaker 1984). 'A person often isn't "of one mind" in accepting a plan or not' (p. 153), he says. This is supposed to be true not just of thinker-planners but also of ordinary people.

While the thing-to-do predication is said to express only a plan, however, it is important for Gibbard that the plan expressed is, as I put it, presumptively rational. He allows that any agent is capable of pro-

nouncing an option as the thing to do, when it fails to cohere with an unquestioned set of background plans and representations, in which case the agent is irrational. He thinks that by resort to the expressive resources linked with the-thing-to-do predicate an agent can come to recognize this sort of failure, revising one or more of the incoherent states. And he suggests that the rational, properly functioning agent will adjust suitably in the light of any such insight, so that the search for insights can serve a self-regulative purpose.

This aspect of Gibbard's view comes out in the fact that he thinks that the introduction of the thing-to-do predicate does more than enable agents to reveal their plans; it also allows them to submit their planning to the test of reason. 'We need a predicate that conveys "to-be-doneness". With it we capture all the power of logic and reasoning that predicates give to language and thought' (p. 13). Thing-to-do predications can be connected up by truth-conditional connectives like 'either … or' or 'if …, then …', to generate rational tests on planning. An example is provided by the reflection that Gibbard ascribes to Sherlock Holmes, as the great detective deliberates about how to escape the fiendish Moriarty. 'Either packing is now the thing to do, or by now it's too late to catch the train anyway … It's not even now too late to catch the train … therefore: Packing is now the thing to do' (p. 42). When Holmes lights on packing as the thing to do, he gives expression to a plan. But, more particularly, he gives expression to a plan that is presumptively rational; should his reasoning prove faulty then we would expect him to revise that plan.

In taking this line, Gibbard rejects the use of the so-called Frege-Geach argument against expressivism, according to which any sentences that can be suitably embedded in contexts like 'either … or' or 'if …, then …' must express beliefs (ch. 3). The states expressed by thing-to-do and okay-to-do predications need not be beliefs, contrary to that supposition; according to Gibbard, they may merely be plans. But will we not have to say, counter-intuitively, that the plan-expressive sentences are at least capable of truth and falsity? Gibbard agrees that we will but he is not worried about this. He allows that despite not expressing beliefs, thing-to-do and okay-to-do predications may be true in a minimal sense of the term; he leaves 'it an open question whether there's a clear and more demanding sense of "truth"' (p. 18).

According to Gibbard's account, then, the expressive resources made available to thinker-planners at this point in the *Aufbau* of their psychology enable them to reason in the ordinary sense; they make it possible for them to achieve something more than the limited, blind

rationality described in section four. Predicating thing-to-do or okay-to-do concepts of options, the thinker-planners can draw conclusions about what plans to form. And, drawing such conclusions, they can regulate themselves by them.

Not only will the expressive resources introduced at this stage enable thinker-planners to reason with themselves about what to do. They will also enable them to reason with one another about such matters. The predicates now available will make it possible for thinker-planners to express their disagreements with each other, and with themselves over time. Planning for someone to do different things—planning to do different things if in his or her shoes—one thinker-planner will say that this is the thing for that person to do, another that that is the thing for the person to do; there will be genuine disagreement between them. They will be in a position to enjoy the advantage, discussed in section three, that Gibbard associates with having access to mutual disagreement and joint deliberation. They will never plan alone.

6. The admission of meta-propositional concepts

Gibbard's view, then, is that although thinker-planners can reason, availing themselves of the concepts of what is the thing to do, or what is okay to do, they need not instantiate any states other than the decisions or plans that might have been there in the absence of such concepts; they need not instantiate any new representations by virtue of exercising the concepts. Why 'think that deciding and thinking what one ought to do are separate activities?' (p. 13). His position keeps two mutually antagonistic claims in play. The capacity of thinker-planners to reason with themselves and others presupposes the availability of potentially regulative predications, marking a break with the blind performance of rational but unreasoning agents. And yet those predications bring no new representations on stream. At one level they change everything, at another they leave everything as it was.

This claim may give us pause. We can certainly agree with Gibbard (p. 5) that the to-be-doneness or okay-to-do-ness of an option is not a property that is there, waiting to be represented, in the world. But why should these properties not be taken as features of the alternative in relation to the agent's background plans and representations: in relation to plans and representations that are taken as fixed and unquestioned? Being the thing to do would consist in being rationally required by such background plans and representations, being okay to do in being rationally permitted by them. Why should finding and proclaim-

ing an option to be the thing to do or okay to do not be a matter of ascribing such plan-relative rationality? In other words, why should it not involve a cognitive state: a higher-order belief in the option's being rationally permitted or rationally required—permitted or required under the combinatorial constraints discussed earlier?

This possibility may seem to be ruled out under the terms of the *Aufbau*, as they have been laid down so far. For no room has yet been explicitly made in Gibbard's story for thinker-planners to form metapropositional beliefs about relations of compossibility between contents and about the associated rationality of holding by certain combinations of contents, whether in planning or indeed representational mode. But at this point it turns out that Gibbard has no objection to such higher-order beliefs about what he calls plan-relative rationality: the rationality of an action or plan of action in relation to certain background representations and plans (p. 93).[1] Given this concession, the answer to the problem raised cannot be that no room has yet been made for thinker-planners to form suitable higher-order representations; such representations have already been admitted.

Allowing that thinker-planners may have beliefs in plan-relative rationality, Gibbard offers a different answer to the claim that saying something is the thing to do is just ascribing plan-relative rationality to it; the case of okay-to-do predications will not be so straightforward but I am happy to concentrate on the example where he is at his strongest. He points out that even a detached observer, with no plan in mind for what the agent should do, can ascribe such plan-relative rationality, and that something else must be involved in finding an option the thing to do for that agent.

Suppose I say that something is the thing for me to do. When I do this, according to Gibbard, I will not stand back like an observer and just register that, taking my background plans and representations as unquestioned, this option is rationally the thing for me to do. Rather, situated at the locus defined by those plans and representations, as it were, I will think or say: this is the thing to do. That predication of the concept will not express a belief to the effect that the action fits rationally with background plans and representations, or anything of the kind; it will express or avow the plan itself.

[1] As suggested in the final section, Gibbard's account of expressing a state of mind as purporting to have it (p. 77) already presupposes the presence of a reflective, self-targeted belief in the thinker-planner. One can hardly purport to have a state when one does not have a belief that one is in that state or, to cover deception, a belief that one might be taken to be in the state.

The expression or avowal may be sensitive to information about the rationality of the plan in the sense that should I learn that the plan does not fit with unquestioned, background plans and representations, then I will drop it; the expression, as we can say, may be the avowal of a plan that is presumptively rational. But there is a crucial difference, in Gibbard's view, between exercising a 'plan-relative' concept in characterizing an option as rational against a certain background, even an unquestioned background, and thinking of it under a 'plan-laden' concept as the thing to do (p. 93).

This should still give us pause. Consistently with the perfectly reasonable observation that saying something is the thing to do is not the same as expressing a belief in the plan-relative rationality of that option, Gibbard might say that it expresses such a belief and does something else as well. The belief that a plan is rational will naturally play a role in generating the plan. How could it be otherwise? Gibbard might hold, then, that saying something is the thing to do expresses a belief that the plan is rational and, given the special plan-generating or plan-laden role of that belief, serves at the same time to express or avow the plan itself. To say that the plan is the thing to do, under this account, will be to express the belief that the plan is rational—rational, inevitably, from one's own point of view—but it will also be to express or avow the plan itself, representing it as a plan endorsed as a generator of action.

This possibility does not cover just the case where I express a belief in the rationality of a plan bearing on what I am to do. It can be extended also to the case where I express a belief in the rationality—again, from my point of view—of a plan bearing on someone else. Suppose I express a belief in the rationality of a plan, if in Caesar's shoes, not to go to the Senate. In doing this, I will indicate that I actually have that plan, counterfactual and fanciful though it may be. For if I endorse it as rational, then I must take it to be the plan I espouse and I will communicate this espousal at the same time as I communicate the belief.

Gibbard does not pay attention to the sort of possibility just rehearsed. Perhaps the reason is that he recognizes the role of meta-propositional representation only as an afterthought when he considers and rejects the suggestion raised in this section. He never explores at length the possibility that it is the appearance of meta-propositional representations—in effect, beliefs in plan-relative rationality—that enables us to make best sense of thing-to-do predications, not the turn to expression on its own. I argue in support of that possibility in the final section.

7. The trappings of realism

Cognitivism takes practical thought to involve cognitive states; an example would be the doctrine that the thing-to-do predicate expresses a plan-generating belief in the plan-relative rationality of an option. Realism would go one step further in asserting that the sort of property ascribed in practical thought really does exist; to go back to our example, there is a bona fide property that constitutes plan-relative rationality. One of the most interesting aspects of Gibbard's discussion is his demonstration that if his non-cognitivist or expressivist picture of thinker-planners is correct, still many apparently realist aspects of practical thought will be vindicated; there will be real properties associated with the application of practical concepts (pp. 88–98).

Consider someone's hyperdecided counterpart, Zeus. This godly counterpart will have settled on what it is okay to do in every conceivable situation for every conceivable agent. He will have registered what naturalistically describable alternatives fall under that concept. But this means that the things that Zeus thinks are okay to do will constitute a naturalistic class, however wild and disjunctive. His distinctions between what is okay to do and what is not okay to do will supervene on naturalistic differences. What it is okay to do by his lights will be 'constituted', as Gibbard himself likes to put it, by a naturalistic property (p. 96). It will be constituted by what realists must take as a real property of things.

Not only will Zeus endorse these claims about supervenience and constitution, of course, so by parity of reasoning will any hyperdecided agent. But this means that all the hyperdecided counterparts of any thinker-planner will endorse those claims. And in that case, according to the account given earlier of rational commitment, any such thinker-planner will be committed to those claims; the claims will be held by all that agent's hyperdecided counterparts. While satisfying expressivism in their practical thought, then, thinker-planners will each be able to acknowledge that corresponding to their thing-to-do or okay-to-do predications, at least if they are consistent, will be a naturalistic property.

While the trappings of realism are thereby preserved, however, this is not realism in any ordinary sense, as indeed Gibbard himself insists. For one thing, the 'direction of explanation', as he says, is wrong (p. 189). The thinker-planners will not think that something is the thing to do, or okay to do, because they think it has a given, naturalistic property. They will think that it must have such a property because of thinking that it is the thing to do or okay to do, where thinking this is merely a matter of deciding on it as something required or permitted.

8. To normative thought in general

The argument about thinker-planners provides the main line of development in Gibbard's book. But the point of the book, of course, is to suggest that our ordinary normative thought operates in a way that parallels or is continuous with the practical thought of thinker-planners. The slogan is: 'Thinking what one ought to do is thinking what to do' (p. 195). In developing the suggestion, there are two key claims to which Gibbard subscribes, one bearing on thin normative concepts, the other on thick.

He claims that the thin concepts of 'ought' or 'good' or 'rational' are plan-laden in a way that is very close to the concept of what to do. Roughly speaking, 'ought', in ordinary parlance, directs an agent to what to do (p. 153), 'good' or 'better' directs an agent to what to prefer (p. 142), and 'rational' directs an agent to how to reason (p. 152). While such terms may be associated in common discussion with presuppositions about the naturalistic features of what is commended—say, assumptions about the sorts of things that can be called good or bad for human beings—'persistent disagreement doesn't raise a question of whether anything is genuinely at issue' (p. 158). Thus, the essential role of any such claim is just to mark out what, from the point of view of the speaker's plans, is the thing to do.

Gibbard's claim about thick concepts is intriguing. Following Geoffrey Sayre McCord (1997), he introduces the idea of a normative kind, though with his own variation. A normative kind is a property that can be invoked, as a matter of common presupposition, in explaining why something is or is not the thing to do, is or is not okay to do. To use a term like 'fair' or 'cruel' of an alternative, then, will be to endorse the associated presuppositions and to point to an explanation—no doubt one that can be overridden—of why that alternative might or might not be okay to do, or the thing to do. The important point about this account, from Gibbard's perspective, is that it allows someone who is an alien to a society, or a radical within it, to reject the presuppositions; this will amount to having plans that those presuppositions cannot help to explain. And those who reject the presuppositions will not then be able to use the terms in the proper normative sense; they will only be able to use them within 'scare quotes' that indicate the distance at which they stand from the community.

In defending these claims about our ordinary use of thin and thick normative concepts, Gibbard establishes to his own satisfaction that we can see ourselves as agents who are fundamentally of a kind with the

thinker-planners he has described. As those thinker-planners allegedly use terms like 'the thing to do' and 'okay to do' in a purely expressivistic way, so we can be represented as using our normative concepts in the same manner. The psychology of the thinker-planners can be taken as our psychology and the semantics appropriate for their practical concepts can be extended also to cover ours.

9. From expressivism to cognitivism

My account of Gibbard's *Aufbau* is faithful to his argument but it is designed to highlight the fact that while he focuses on the turn to expression as crucial in making sense of thing-to-do and okay-to-do predications, he recognizes a distinct development—the capacity for meta-propositional representation—that promises a more straightforward, cognitivist account of such predications. This account would escape the tension that arises between his claim that expressive resources are deeply important, making it possible for thinker-planners to reason with themselves and with one another, and his claim that those resources bring no new representations on stream. It would support the view that thing-to-do predications express plan-generating beliefs in the plan-relative rationality of certain options; here, as in the discussion in section six, I ignore okay-to-do predications, which are going to be trickier for Gibbard to handle.

We can develop this argument by starting with representing rather than planning. It is one thing for a subject, perhaps just a piece of artificial intelligence, to be so constituted that it updates its beliefs inferentially, at least within certain limits; given a standing belief that if p, then q, it more or less automatically goes to the belief that q, on being primed by incoming evidence to believe that p. It is quite another thing for that subject to be able to form meta-propositional beliefs about its existing belief-contents—say, the propositions 'if p then q', and 'p'—to the effect that they entail that q, and to be disposed to adjust its own performance in light of such a meta-belief: to be disposed in such a case, assuming its existing beliefs are not up for question, to form the belief that q.

The latter possibility is familiar and straightforward. The agent takes certain propositions to be unquestioned, such as 'if p then q' and 'p', he or she goes on to recognize that those propositions entail that q, and then the belief in the entailment plays a special role: it elicits or reinforces the belief that q in the agent, playing the part of a conclusion-generating belief. Agents may be said to regulate their own perform-

ance in these regards just so far as they intentionally inquire into such entailments, in the expectation that their emerging, meta-propositional beliefs about entailments will have a conclusion-generating effect. A sign of such self-regulation will be the fact that they will not just form a representation expressed as 'q'; they will be in a state that they can express by saying 'So q' or 'Therefore q'.

That the belief in an entailment can have such an effect is not surprising. Were a subject able to develop entailment beliefs without any tendency to update according to those beliefs, then he or she would suffer from a barely intelligible failure of coherence and rationality (Smith 1994). To believe that p, that if p, then q, and that the truth of those propositions entails the truth of 'q', is to incur a rational obligation, assuming no other belief-change is on the cards, to believe that q. Not to believe that q in the presence of those other beliefs represents an even more egregious version of the failure involved in believing that p, that if p, then q, but not that q; and by almost all accounts, rational agents must be assumed to be organized so that even the lesser failure is generally avoided.

If this line of thought is sound, then it suggests that there is a natural candidate for the state that might be expressed by describing a proposition as 'the thing to accept', to introduce a predicate analogous to 'the thing to do'. That state will be the belief that the proposition is rationally supported by propositional or perceptual evidence in which the speaker has suitable confidence. Such a state ascribes evidence-relative rationality to the proposition, yet does so while playing a special role. Not only does it track the rationality of the proposition in relation to the evidence; it serves in the rational or coherent agent to elicit or reinforce a belief in that proposition. To say then that a proposition is the thing to accept will be to express the belief that it is rational but it will also be, inevitably, to express a belief in that proposition itself.

This analogy suggests a parallel for the case of practical reasoning. While remaining faithful to other aspects of Gibbard's argument, we might cast the situation with his thinker-planners on the lines gestured at earlier. These agents can form a meta-propositional belief about the plan-relative rationality of doing something, as they can form a belief about the evidence-relative rationality of drawing a certain conclusion. As their belief in evidence-relative rationality will have a special role to play in prompting them to draw the relevant conclusion, so their belief in plan-relative rationality will have a similar role to play in prompting them to decide on the relevant option. And as the expression of the former belief is bound also to express the drawing of the conclusion

itself, so the expression of the latter is bound also to express the formation of the plan.

There is nothing surprising in the fact that the belief in the evidence-relative rationality can play a conclusion-generating role. And by parity of reasoning there is nothing surprising in the fact that a belief in the plan-relative rationality of an option can play a plan-generating role: it will be a condition of coherence or rationality that the belief should generally have that sort of effect. This being so, I can be taken to use the phrase 'the-thing-to-do' to express a belief that an option is rational and, at one and the same time, to express a plan that favors that option.

What should we say about the case where I have plans for another person—plans for what to do if in the other person's shoes—and beliefs about what is rational relative to those unquestioned plans? It may seem that while holding such beliefs, and while being wholly rational myself, I need not let the beliefs play the role of forming a plan for what the other decides: I may remain quite detached. But actually this is not so. It is one thing to believe that an option is rational relative to the background plans of the other person, where I need not endorse those. It is quite another to believe that it is rational relative to my background, unquestioned plans for the other person. And if I believe it is rational relative to my plans for the other person—assuming that we go along with Gibbard in allowing for such plans—then I will naturally be disposed to act appropriately if in the shoes of that other person.

The cognitivist version of thing-to-do predications, or indeed of okay-to-do predications, is unlikely to recommend itself as a full theory of moral thought. If we focus on what is rationally required of us—required in virtue of states to which we are already committed—then we are naturally led to ask about whether anything is required in other ways. Apart from considering what we are rationally obliged to believe in view of believing other things—other things that are taken as unquestioned—we naturally ask the distinct question as to what we have most reason overall to believe: most reason to believe in light of how things are. And similarly, apart from considering what we are rationally obliged to plan and do in view of our unquestioned, background plans and representations, we naturally raise the distinct question as to what we have most reason overall to do (Broome 2004).

This feature will make the cognitivist version of his view unattractive to Gibbard, for no similar limitation imposes itself, if we think of thing-to-do predications in purely expressive terms. But still, there are other reasons why the cognitivist picture should retain some appeal for him. First, it would explain nicely the relation between plan-laden beliefs

and beliefs in plan-relative rationality, rather than have them appear as quite different states of mind; on this picture, the plan-laden belief that something is the thing to do ascribes the same property as the plan-relative counterpart but it is recruited to a special role. Second, the cognitivist line would explain why the plan-laden belief about what to do oneself might remain unchanged, as a plan-laden belief, and yet fail for contingent reasons—say, weakness of will—to discharge its normal role; there will be no need in such a case to resort to the view that the subject is literally of many minds, with only one mind truly planning. And, third, this view would make it entirely unsurprising that the sentences in which thinker-planners debate about what is the thing to do behave in a regular, truth-conditional manner; there will be no need for special efforts to explain the fact that they perform as if they were reports of facts: they are reports of facts, according to the view, but reports from a special, plan-generating perspective.

Apart from these particular attractions, the cognitivism I recommend would also enable Gibbard to stand in the sort of place that he clearly finds appealing. While being naturalistic in its underlying ontology—as Gibbard himself is a naturalist (p. 191)—it would deny, as he denies (p. 103), that practical and normative statements and thoughts ascribe natural properties 'as such'. They ascribe them only to the extent that the properties play certain roles in our reasoning and motivation: roles that must include the planning role, even if they also have other aspects (Jackson 1992, Jackson and Pettit 1995, Jackson 1998, Pettit 2001).

10. A problem for expressivism

Apart from having these attractions, the cognitivist view sketched would avoid a general problem that Gibbard must face if he sticks with his expressivist alternative (Jackson and Pettit 1998, see too Jackson and Pettit 2003, in reply to Smith and Stoljar 2003). He does not address the problem as it arises for his account, referring to it only in passing (p. 76), but I think that it represents a major hurdle for his approach.

Like everyone else, Gibbard assumes that a phrase like 'the thing to do' will be used in the normal, voluntaristic manner of a linguistic sign. It will be intentionally employed to convey the presence of whatever the phrase is taken by definition or convention to be a sign of: the fact, as under Gibbard's own account, that an option is the content of a plan. Thus an agent's self-targeted use of 'the thing to do' will 'purport' (p. 77)

to communicate a plan to do it; presumptively, a plan that rationally fits with the agent's background plans and representations.

But how is the agent supposed to keep track of when it is proper to use the phrase 'the thing to do', and when it is not proper to use it? How is the employment of the phrase to be guided? It cannot be guided just by the existence of the relevant plan. How could that plan prompt an utterance of the words 'that is the thing for me to do', without my being sensitive to the fact that I have the plan: without a belief, in effect, that I have the plan? The plan might involuntarily elicit the phrase, as boredom elicits a yawn, or amusement a smile. But if the utterance of the sentence is to be a voluntary, intentional act, then Gibbard must allow that it is subject to the guidance, however implicit or virtual, of a belief that there is a plan present that it serves to express.

Gibbard cannot easily make room for the role of such an utterance-guiding belief, whereas the cognitivist alternative I have been sketching accommodates it easily. On that story, the sentence 'this is the thing to do' reports the content of the belief that this is an option that is rationally required relative to one's unquestioned plans and representations. And such a belief can provide the required guidance in the use of the thing-to-do phrase.

Under this alternative story, a sentence characterizing an option as the thing to do will operate just like a sentence characterizing an object as the largest one around. It will express the belief that the option is plan-relatively rational, reporting how things are according to that belief in words that are believed to be expressively appropriate. And it will do this intentionally, out of a desire to communicate the fact that the option is plan-relatively rational and, by implication, the fact that one plans to choose it.[2] Thus there will be no problem about how the use of the sentence is guided; it will be guided by the usual combination of belief and desire that we expect to find at the origin of any voluntary act.

For this reason, and for the reasons rehearsed earlier, I believe that the cognitivism at which I have been gesturing is more attractive than Gibbard's expressivism. But even if such cognitivism is more appealing than Gibbard's own expressivist doctrine, I do not think that this reflects badly on the book. On the contrary, the very fact that a non-expressivist can find so much to applaud in this book is a tribute to its richness.

[2] Typically, of course, one will also communicate, by conversational implicature, the fact that one is willing to have others rely on one's enacting the plan. Why would one report on a plan, if not to communicate such willingness? The point of the utterance is hardly to inform hearers about the state of one's plans, without inviting any reliance on one's behaviour; only a therapist would find such information conversationally relevant.

It is a treasure trove of insight and argument and no one, however unconvinced about its bottom line, will come away disappointed.[3]

University Center for Human Values PHILIP PETTIT
Louis Marx Hall
Princeton University
Princeton, NJ 08544
USA

References

Blackburn, S. 1984: *Spreading the Word*. Oxford: Oxford University Press.

Broome, J. 2004: 'Reasons', in Wallace, Smith, Scheffler, and Pettit 2004.

Carroll, L. 1895: 'What the Tortoise said to Achilles'. *Mind*, 4, pp. 278–80.

Jackson, F. 1992: 'Critical Notice, S. Hurley Natural Reasons'. *Australasian Journal of Philosophy*, 70, pp. 475–87.

Jackson, F. 1998: *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, F. and P. Pettit 1995: 'Moral Functionalism and Moral Motivation'. *Philosophical Quarterly*, 45, pp. 20–40, reprinted in Jackson, Pettit, and Smith 2004.

Jackson, F. and P. Pettit 1998: 'A Question for Expressivists'. *Analysis*, 58, pp. 239–51.

Jackson, F. and P. Pettit 2003: 'Locke, Expressivism, Conditionals'. *Analysis*, 63, pp. 86–92.

Jackson, F., P. Pettit, and M. Smith 2004: *Mind, Morality and Explanation*. Oxford: Oxford University Press.

Leiter, B. (ed.) 2000: *Objectivity in Law and Morals*. Cambridge: Cambridge University Press.

Pettit, P. 2000: 'Embracing Objectivity in Ethics', in Leiter 2000.

Pettit, P. 2002: *Rules, Reasons, and Norms: Selected Essays*. Oxford: Oxford University Press.

Sayre-McCord, G. 1997: '"Good" on Twin Earth'. *Philosophical Issues (*supplement to *Noûs)*, 8, pp. 267–92.

Smith, M. 1994: *The Moral Problem*. Oxford: Blackwell.

[3] I was helped in preparing this paper by comments received when it was presented in a workshop at the European Conference for Analytical Philosophy, Lisbon, 2005 and at a seminar in the Australian National University; by conversations with John Broome, Frank Jackson, Gideon Rosen, Michael Smith and Daniel Stoljar; and by comments from an anonymous referee. Allan Gibbard helped enormously by clarifying some points of interpretation.

Smith, M. and D. Stoljar 2003: 'Is There a Lockean Argument against Expressivism?' *Analysis*, 63, pp. 76–86.
Stalnaker, R. C. 1984: *Inquiry*. Cambridge, MA: MIT Press.
Wallace, R. J., M. Smith, S. Scheffler, and P. Pettit (eds) 2004: *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press.

