IN SEARCH OF OBJECTIVE AGENT RATIONALITY

Daniel Laurier

Département de philosophie
Université de Montréal
C.P. 6128, Succ. Centre-ville
Montréal, Québec H3C 3J7
Daniel.laurier@umontreal.ca

The purpose of this paper is to offer an account of what an agent's being rational to do or think something might amount to, which doesn't reduce to saying that it consists in this agent's doing or thinking something that is rational for him. In the first section, I call attention to the fact that such a distinction between agent rationality and action or belief rationality is widely admitted, I reject the idea that it could be interpreted as a distinction between the rationality of tokens and the rationality of types, and I suggest one natural way in which a notion of objective agent rationality could be informally characterized in terms of action or belief rationality. But this first, rough, characterization depends on further uses of the notion of rationality which I try to make sense of in the second section, at least in a preliminary way.   The burden of this second section is then to determine whether the intuition behind the informal proposal introduced in the first can be substantiated, i.e., whether a substantial and coherent notion of agent rationality can be worked out, and at what cost. In the concluding section, I try to "deflate" some of the worries that could be raised by the account of agent rationality I end up with.

1. Agent vs Action/Belief Rationality

One of the most basic uses of the concept of rationality seems to be that in which we commonly say that it is (or would be) rational/irrational for a certain agent, at a certain time, to act or think in a certain way[1], that is, to do this or that, or to have this or that propositional attitude. This suggests that it is primarily singular actions (or doings) and thoughts (or intentional states) that can be rational/irrational. In what follows, however, I am mainly concerned with belief and action, with the hope that what I am going to say could be generalized to other (and especially conative) intentional states[2]. When we say

---

[1] "Thinking" is here understood in the largest sense, including feelings and emotions, insofar as these may have intentional content.
[2] A question I won't be able to discuss in detail here, and which raises a number of controversial issues.

that agents themselves, or certain dispositions to act or believe in certain ways, are rational/irrational, this can usually (or often) be understood in a derivative way, as meaning that a certain agent does or believes what it is rational/irrational for him to do or believe (or that he is capable or disposed, either generally or at some particular time, to do or believe what it is rational/irrational for him to do or believe), or that certain dispositions are dispositions to do or believe what it is rational/irrational to do or believe for a certain agent (or an agent of a certain kind)[3].

However, there seems to be another way to understand rationality as applied to agents. For it is not obvious that "Agent S is being rational in doing or believing X" is always equivalent to "It is rational for S to do or believe X"; this is because it would (sometimes) seem to make sense to wonder whether S is rational *in* doing or believing what it is rational for him to do or believe, i.e. to wonder whether S is lead to do or believe what it is rational for him to do or believe by going through what might be called a "rational" process (from what might be called a "rational" basis). In other words, an agent might have good reasons to do or believe what he does or believes, and yet not do or believe it for good reasons. Some such distinction between agent rationality and action/belief rationality seems to be accepted by many authors (e.g. Davidson, Firth, Foley, Audi, Alston, Goldman, Bratman, *inter alia[4]*). But what kind of distinction is that, and which notion of rationality (if any) should be taken as basic? It would seem hard to answer these questions without saying more about the content of rationality judgements.

So far, I have suggested that a basic rationality judgement is of the form "It is rational for agent S (at t) to do or believe X", but that this is somehow ambiguous as between

---

3 No doubt, there are other kinds of rationality judgements, as for example when we say that certain rules (maxims or principles) are rational/irrational. But this I understand as meaning that successful applications of the rules would lead a certain agent to do or think what it is rational/irrational for him to do or think. This doesn't imply that I am rejecting any attempt to explain what a rational action or belief is by saying that it is one that conforms to certain rules, but only that if one explains the basic notion of rationality in this way, one thereby deprives oneself of the possibility of raising the question of the rationality/irrationality of these rules themselves (which are then "constitutive" of rationality). It could also be helpful to point out that nothing in what I have said prevents one from saying that it is rational/irrational for a certain agent to acquire (try to acquire) or get rid of (try to get rid of) certain dispositions to act or believe in certain ways, since these are just further actions.

4 Much of what follows can be seen as an attempt to elucidate and generalize a distinction which (as far as I know) was introduced by Firth (1978) and seems to have been admitted by many epistemologists. This is a distinction between two kinds of epistemic justification, which he calls "propositional" and "doxastic", respectively. I depart from Firth and these epistemologists mainly in that (i) I focus on rationality, rather than justification, (ii) I am not specifically concerned with epistemic rationality, and (iii) I think the distinction extends to most kinds of intentional attitudes and actions. Given these interests, it would have been odd to retain Firth's terminology.

(1) Action or belief X is rational for S (at t)

and

(2) S is being rational in doing or believing X (at t),

and that (1) does not imply (2). For one thing, (1) doesn't imply that S actually does/believes X; but more importantly, it doesn't imply that S is rational in doing/believing X, even together with the assumption that S actually does/believes X, which suggests that (1) could not be understood as saying that S would be rational if he were to do or believe X. It is tempting to understand[5] the distinction as one between the rationality of types and the rationality of tokens (or perhaps more aptly, tokenings), so that the ambiguity would be between

(1') Action or belief type X is rational for S (at t)

and

(2') Action or belief token X is rational for S at (t).

One would then have to hold that if some action/belief token is rational for S, then S is rational in doing/believing it. On the other hand, one could allow that some action/belief type is rational for S, while S is not rational in doing/believing some token of that type. There seems to be some truth behind this intuition, but perhaps this is only due to the fact that (since tokens, but not types, are somehow located in spacetime and can enter into causal relations) by talking of types and tokens, one is calling attention to the fact that rationality is sometimes taken as a purely "abstract", "teleological" matter, while at other times, it is taken (at least in part) as a causal-historical (or procedural) matter.

To say that some action or belief (type) X is rational for some agent S, is (very often, though by no means always, as will be made clear later) to say that it entertains a certain relation to S's (actual) intentional states. But the nature of the relevant relation may vary widely, for in saying that action or belief (type) X is rational for S, given the set T of all intentional states of S, one can mean either that (given T) X is at least as "good" as any other alternative (e.g. it maximises his utility function), or that S (or some other agent) *judges* that (given T) X is at least as "good" as any other alternative, or that S (or some other agent) *would judge* that (given T) X is at least as "good" as any other alternative, if he were to reflect in a certain way for a certain time. Foley (1987), for one, takes this (among

---

5 Following Audi (1985) and Kvanvig and Menzel (1990), among others.

other things) to indicate that there are countless perspectives from which the rationality of an action or belief (type) can be evaluated[6]. I have no quarrel with that; but what seems significant is that the third kind of case makes the rationality of an action/belief type relative, not only to the agent's mental states, but also to some set of dispositions to reflect or reason, which may be the agent's own or someone else's. Among these various possible perspectives, some are worth singling out for special consideration, i.e. those where, in saying that action/belief type X is rational for S (at t), one is asserting either that

> (3) given the set T of all intentional states of S (at t), action/belief type X is among the best alternatives (at t),

or that

> (4) S judges that, given the set T of all his intentional states (at t), action/belief type X is among the best alternatives (at t),

or that

> (5) if he were to reflect or deliberate carefully about it for a certain time, S would judge that, given the set T of all his intentional states (at t), action/belief type X is among the best alternatives (at t).

It doesn't matter here, if we are unable to say for how long S is supposed to reflect or deliberate, or what is to count as careful reflection or deliberation. The point about (5) is that this kind of judgement makes the rationality of an action/belief type depend on some dispositions of the agent and not only on his actual intentional states and what they entail (though of course, in a sense, these dispositions to reach a certain verdict could qualify as mental states). Let us call the first perspective the objective perspective, the second the subjective (or actualized) perspective, and the third the subjunctive (or idealized) perspective. It seems intuitively clear that although the subjunctive perspective is somewhere between the purely objective and the purely subjective, it is "conceptually" more complex than the subjective perspective, which is itself "conceptually" more complex than the objective perspective.

A further thing to make clear is whether, in making such judgements as (3)-(5), one means to be talking about an action/belief type which is actually performed/instantiated or

---

6 As a matter of fact, Foley allows for much more flexibility than I am doing here, since in his views only the beliefs of the agent, and not the set of all his intentional states, are to be held constant in making rationality judgements; so that in

only about one which could be performed/instantiated. It seems best to read them in the second way, since it would always be possible to express the first reading in some other way, by saying that some action/belief token is of a type X which is among the best alternatives (or simply by adding that the agent performs an action of that type or has a belief of that type), or that some action/belief token is of a type that S judges or would judge to be among the best alternatives. This brings it about that[7] the distinction between action/belief rationality and agent rationality doesn't clearly reduce to a distinction between the rationality of types and the rationality of tokens, for when some action/belief token is of a type which is rational for S, it is always permissible to say that *it* is rational for S in virtue of being (or *qua* token) of such a type, even though S himself is not rational in doing an action of that type or having a belief of that type[8]. Furthermore, when some action/belief token is not rational, it must be in virtue of some of its features, i.e., in virtue of its belonging to some type (which then fails to be rational). But if so, how are we to characterize agent rationality?

Before returning to this question, let us make it clear that since an agent's total mental state (as I will sometimes call the *set* T of all his intentional states) includes all his beliefs, desires or preferences, the rationality judgements here in question are "all things considered" (or global) rationality judgements. But of course, nothing prevents one from assessing the rationality of some action/belief relative to some subset of the agent's intentional states (especially his desires, goals or values), or even relative to any set of intentional states. What seems important, is that when doing this, the relevant set of states is still taken as an hypothetical *total* set, in the sense that in saying that X is rational for S relative to some subset G of his intentional states, one is in effect saying that X would be rational for S if G were to exhaust his intentional states (with no presumption that it would remain so if G were changed in any way).

From what has been said, it would seem that action/belief rationality is a matter of "fit" between some (actual or possible) action/belief and an agent's intentional states, while agent rationality is a matter of "fit" between some actual action/belief and the way in which

---

making a rationality judgement, one could for example be evaluating how well various alternatives contribute to reaching some subset of the agent's (or one's own) goals or values.

7 Contrary to what is suggested in Kvanvig and Menzel (1990).

8 In any case, it would hardly make sense to say of types that they are rational/irrational, if that didn't carry the conclusion that their tokens are as well, if only as belonging to these types.

the agent comes (or came) to perform that action or to have that belief. This brings us back to the idea that to say that an agent is rational *in* doing/believing X is fundamentally to make a judgement concerning the causal history or etiology of some action/belief of his, i.e. some *actual* process or sequence of events. But one must be careful here, for in asking whether the agent is rational in believing (as opposed to doing) something, one may be understood either as asking whether the agent is (or was) rational in *forming* the given belief or as asking whether the agent is rational in *continuing* to hold the belief. As it is often noted, someone may, at some point in time, be rational in having a belief even though he was irrational in acquiring or forming this belief. The same thing obviously could not happen in the case of actions, since they cease to exist once they have been done[9]. It is important to be clear about the fact that when I ask at t whether some agent is rational in doing/believing X, I mean to be asking whether he is rational *at t*, in doing X or in forming the belief X or (if he already held the belief) in persisting in the belief X. Insofar as "forming a belief" can be understood as a special case of "doing something"[10], it will be convenient to systematically read the question whether an agent is rational in believing something as a question concerning the rationality of his continuing to believe it (as opposed to the rationality of his coming to believe it)[11].

Once this has been seen, it is no longer clear in what sense it can be maintained that to say that an agent is rational in believing something is to make a judgement concerning the causal history or etiology of his belief. To say this is of course to make a judgement concerning the relation between this belief and its actual ground or basis, and it is natural to take the basis of a belief to be something that at least causally sustains it (or causally explains its *continued* existence). So even though it may be slightly misleading, in such a case, to talk of a "causal process", I will stick to that usage at least for the time being (if the need arises, I could talk of "causal-sustaining" and "causal-producing" processes).

---

9 This may need some qualification, for it could happen that the agent starts doing something for certain reasons, and at some point in the course of doing it, keeps on doing it for some other reasons. But then it is not the case that the action has been completely done for some set of reasons, and then maintained for some other set of reasons, as might well happen in the case of belief.

10 This raises a number of interesting questions into which I cannot enter here, such as whether "forming a belief" can always or normally be taken as a kind of intentional action, whether it has to be taken in this way for it to be able to be done "for a reason", whether to take it in this way requires that it be under the control of the will (and to what extent), etc. My strongest intuition on these matters, is that forming a belief is often (though not always) done for reasons, and if it turns out that it is not an intentional action, I should conclude that not only intentional actions can be done for reasons.

11 This is not to deny that, in some sense, "continuing to believe X" is also something that agents can do, and sometimes do for reasons (even if they cannot do it "at will").

If saying that S is rational in doing/believing X is only to make a judgement concerning the way in which S's action or belief is caused or causally sustained, then S could be rational in doing/believing X even when X is not rational for S (given his total mental state), provided his doing/believing X results from what might be called a "rational" process (in the extended sense I have just alluded to) starting from what might be called a "rational" basis.   But there is also a stronger sense of the phrase, according to which it cannot be the case that S is rational in doing/believing something that is not rational for S given his total mental state. In this stronger sense, one might suggest that S is rational in doing/believing X if *and only if* X is rational for S given his total mental state *and* S's doing/believing X results from a rational process starting from a rational basis.

But at first sight, it seems this cannot yet be the whole story, for even when these two conditions hold, one might still complain that S is irrational in doing/believing X, because even though X is rational for S given his total mental state, either the latter is not rational (*tout court*) in the first place or S's being in this total state is not rational. In other words, it is *prima facie* plausible to suggest that, as a first rough approximation:

> (6) S is rational (at t) in doing/believing X iff (i) S's doing/believing X results from (or rests on) a "rational" process starting from a "rational" basis, (ii) X is rational for S given the set T of all his intentional states (at t), (iii) T is itself rational for S and (iv) S is rational (at t) in being in T.

But insofar as X must here be taken as belonging to T (since only if S does/believes X can he be rational in doing/believing it), this is obviously circular and should be rephrased so as to make reference not to T itself, but to the set T-minus-X of all intentional states of the agent that remain when X has been removed[12]. Furthermore, since it has been posited that S cannot be rational in doing/believing X unless T-minus-X is rational for him, it has to be the case that S cannot be rational in being in T-minus-X unless T-minus-X is rational for

---

[12] T-minus-X may or may not be the set of all intentional states of S *other than X*, depending on whether beliefs or intentional states can or cannot come and go one at a time. If it turns out that they cannot, it would have to be understood that removing X from T is not a simple set-theoretic operation; but as far as I can see, the ensuing complications would not be insuperable. Furthermore, should it turn out that intentional states always or sometimes come in bundles, one would still need to be able to tell whether the states in any such bundle are rational relative to each other (or else take these bundles to be the relevant units), which suggests that it must be possible to make sense of the claim that some X is rational given the set of all other intentional states of the agent, even when this set contains states that the agent could not be in without being in X. Nothing, therefore, seems to prevent one from taking T-minus-X to be the set of all intentional states of S, except X.

him. It follows that condition (iv) implies condition (iii), and that our first approximation boils down to:

> (7) S is rational (at t) in doing/believing X iff (i) S's doing/believing X results from a rational process starting from a rational basis, (ii) X is rational for S given the set T-minus-X (obtained by removing X from the set of all his intentional states), and (iii) S is rational in being in T-minus-X.

It is of course to be expected that the "basis" of any process leading to or sustaining some action/belief (insofar as it is an "intentional" basis) has to be included in T-minus-X, but as will appear below, this doesn't mean that the basis of any such process will be rational whenever the agent's total state is. Obviously, when S's doing/believing X is based on *nothing less* than T-minus-X, his doing/believing X will turn out to be rational just in case it results from T-minus-X by a rational process and S is rational in being in T-minus-X; for this will guarantee both that X is rational for S given T-minus-X and that the basis of the relevant process is a rational one. But otherwise, it seems it could be the case that S's doing/believing X results from a "rational" basis by a "rational process" without X's being rational for S given T-minus-X (and this is why the latter condition has to be made explicit).

2. Objective Agent Rationality

The basic intuition, then, is that agent rationality (in the strongest sense anyway) depends not only on the rationality of the process leading the agent to do or to continue to believe something, but also on both that of the starting point of this process (i.e., on the "rationality" of the intentional states on which his action/belief is ultimately based *and* of his being in these states) and that of its result. If this is even roughly on the right track, it means that a substantial characterization of agent rationality would have to rest on an account of what it is for a "process" to be rational, and of what it is for a set of intentional states to be rational *tout court*, i.e. in some *absolute* sense (and not only of what it is for an intentional state to be rational *given* some set or system of intentional states)[13]. But before we could even try to give an account of process and absolute rationality, we have to make

---

13 This clearly suggests that I'm taking it for granted that the notion of agent rationality can be adequately explained (if at all) on the basis of an explanation of action/belief rationality and "process" rationality. But I don't thereby commit myself to the view that action/belief rationality could not, in turn, be explained in terms of agent rationality, though I won't enter into this question here (but see, e.g. Kvanvig and Menzel (1990)).

sense of the idea that a "process" might be rational/irrational, and that a set of intentional states (and thus a single action/belief) might be rational/irrational *tout court*.

As far as the last question is concerned, the two basic options would seem to be to espouse a form of coherentism or a form of foundationalism. The first view is that the rationality/irrationality of any intentional state always depends only on its relation to the agent's other intentional states, while the other is that the rationality/irrationality of *some* intentional states does not depend (at all, or only) on their relations to any other intentional state of the agent but on their relations to something else (or to nothing at all, these being somehow "intrinsically" rational[14]). Now I don't see what can be gained in saying that some intentional state might be rational for S (if only in part) in virtue of some relation that it has to something which is not an intentional state of S, for it would seem that no fact or event could make anything rational for S unless S recognizes it (and recognition is an intentional state); and in any case one would then be using a totally different concept of rationality.

It has been suggested, for example (by Jonathan Dancy), that if I have borrowed a book from Pierre, this may be sufficient to make it rational for me (ceteris paribus) to return the book to him (perhaps together with some general principle such as "one ought to return to its owner what one has borrowed from him"). There are four points to be made in reply to such a contention. First, it would seem odd (to me at least) to say that it was *rational* for me to return the book to Pierre if, for example, I have completely forgotten having borrowed it from him, and perhaps even come to believe that it was mine. Second, admitting this doesn't prevent one from saying that my returning the book was "appropriate", or "morally good", or perhaps even "justified", in some other sense. In other words, it is open to one to hold that external facts might contribute to determine when something satisfies certain (moral, epistemic, etc.) norms, while denying that these are norms of *rationality*[15]. Thirdly, (given that reasons are defeasible) it is to be feared that if one accepts that *facts* (and not only intentional states) can make it rational for someone to do/believe X, this would have the absurd consequence that an action/belief can be rational only given the totality of facts (i.e., a complete history of the world). Finally, this would

---

14 So, on this way of contrasting coherentism and foundationalism, coherentism is a form of internalism, while foundationalism can be either internalist or externalist.

15 Though I admit this is a price that many would not be prepared to pay.

offend my intuition that only something which can be said to be rational/irrational can make something else rational/irrational, and *facts* as such can hardly be described as rational/irrational.

   As for the idea that some intentional states might be intrinsically rational (if only for one agent or another), it would certainly call for further clarification, but *prima facie,* there would seem to be several ways of making sense of it, if only by saying that an intentional state is "intrinsically" rational for S just in case it would be rational for S, no matter what other intentional states he is in. In thinking about what intentional states might be intrinsically rational for S, one is likely to think of such "tautological" states as believing that if p then p, which seems to be somehow "required" by reason (and also to fit the suggested gloss). However, it could hardly be the case that in saying that some action/belief is rational *tout court* for S it is always meant either that it is "intrinsically" required by reason, or that it is rationally required by intrinsically rational states of S. That would be overly restrictive; and it is natural to allow that something that is not rationally required may not be rationally prohibited either, and thus can count as rational in the sense of being rationally "permitted". The idea that a rational action/belief is one that is "permitted" rather than one that is "required" is already implicit in the claim that to say that an action/belief is rational for S given T is tantamount to saying that given T, it is at least as "good" as any other alternative. For more than one action/belief may be as good as any other alternative, and it is normally not the case that each of them is "required". The claim that some intentional states are "intrinsically" rational could be understood in the same (negative) way, as saying that they are not "intrinsically" prohibited; i.e., as saying not that they would be rationally permitted no matter what other intentional states the agent would be in (for then, again, only "tautological" states would fill the bill), but that they would be rationally permitted by at least some other intentional states the agent could be in. But so understood, it doesn't have much bite, for all intentional states would turn out to be "intrinsically" rational for S, except those that are prohibited for S no matter what other intentional states he is in (and are thus "internally" defective). In any case, it seems we have to take account of the fact that an action/belief may be either rationally *required* (R-rational), or rationally *permitted* (P-rational) or rationally *prohibited* (irrational). And when P-rationality is taken as basic, there would not seem to be much to choose between

coherentism and foundationalism (as far as we are talking about action/belief rationality, and not about agent or process rationality). With this in mind, let us return to our two "basic options".

On the coherentist view, one might say that a (total) set or system T of intentional states counts as P-rational (*tout court*) for S if and only if every X belonging to it is P-rational for S given T-minus-X[16]. On the foundationalist view, it would seem natural to suggest that T counts as P-rational for S if and only if there is some subset C of T such that each member of T not contained in C is P-rational for S given C, and each member of C is (i) P-rational for S given the set of all other members of C (more precisely, given the set obtained by removing it from C) and (ii) P-rational for S *in some further sense*, as for example in virtue of entertaining (or not entertaining) some relation to something other than S's intentional states, or in virtue of having some "intrinsic" quality. If there is no such "further sense" available, then condition (ii) is empty. As I have already remarked, I don't see how any relation to anything other than an intentional state can make anything rational or irrational. I don't see what kind of "intrinsic" quality can do the trick either. It is granted that it does make sense to say that an intentional state is intrinsically P-rational just in case it would be P-rational for S, given at least some set of intentional states he could be in. But if there are such states, they are already taken care of by condition (i), and there is no need for condition (ii). Furthermore, it is dubious that "intrinsic P-rationality" makes appeal to any "intrinsic quality", when it is characterized in this way. However that may be, what then remains of the foundationalist view is still not acceptable, for the set T of S's intentional states could then count as P-rational even though it contains states that "rationally" exclude one another, for such intentional states could each be rationally permitted by the set C of "basic" intentional states.

The foregoing remarks concerned the rationality of (sets or systems of) intentional states. However, things look somewhat different when we turn to agent or process rationality. For consider. On the kind of coherence view we are contemplating, the P-rationality of a set of intentional states depends only on the (abstract) relations its members have among themselves. But when S actually is in some intentional state, his being in that state is either based on other intentional states of his, or on no other intentional state (of his)

at all; but insofar as the basing relation is at least partly causal, it cannot be the case that all of S's intentional states are mutually based on each other[17]. This suggests that one could hardly *stipulate* that, if all members of T are P-rational for S given all other members of T, then S is P-rational in being in each of these states, *except* when no member of T is based on any others, i.e. when all members of T are exclusively based on non intentional (external) factors. In other words, when it comes to agent or process rationality, it would seem there is no other choice than to accept some form of foundationalism.

Let us now turn to "process" rationality. At first sight, it would seem that for it to make sense to say that some (actual) process or sequence of events is rational/irrational, it would have to consist in steps or stages that can be described in intentional terms (i.e., as being themselves either actions or intentional states) such as a chain of reasoning, since to reason is (in part at least) to make choices and inferences, and (*prima facie*) these are (mental) actions. However, I suspect that taking this line threatens to lead to serious trouble. So I think it best to start with the idea that the steps leading from some intentional states to some action/belief are purely causal steps, however difficult it might be (and *I do* find it nearly impossible[18]) to conceive of such purely causal links as being "rational/irrational".

I suggest that we understand the claim that S's doing/believing X is based on some set G of intentional states of S as implying not only that S's doing/believing X is caused or causally sustained by members of G, but also that all members of G are (partial) causes of S's doing/believing X. This means that if G is included in some other set F of intentional states of S, then S's doing/believing X could be based on G without thereby being based on F (though it could also be based on F, for if S's doing/believing X is based on some set of intentional states, it is also based on all subsets of this set, as I am using the term).

To find the process leading from G to X, first find the members of G which are not (partially) caused by any other member of G. These will be the starting points of the

relevant process and form its *proper* basis (they might be called G-basic states since they play the part of basic intentional states, relative to G). Then find, for each of these states, the intentional states of S (if any) which it immediately causes (or causally sustains) *and* are themselves partial causes of S's doing/believing X. Do the same for each of these intentional states, and repeat until only states which are immediate partial causes of X are found. The set G# obtained by removing X from the set of all intentional states involved in this process will constitute its *total* basis. I submit, as a first approximation, that

> (8) the *process* leading from G to X will be P-rational for S iff (i) X is P-rational for S given G# and (ii) each member of G# is P-rational for S given the largest subset of G# on which it is based (i.e., the set containing all and only the members of G# that are partial causes of the given state).

This calls for some comments and explanations. It might be thought that, since a process (even in our extended sense) is nothing but a set of paths or sequences of steps (each consisting of a direct or immediate causal link), it would be enough to require that each step be P-rational. It is fairly obvious that a single step from some intentional state Y to some action/belief X should be P-rational iff X is P-rational given Y (more strictly speaking, its unit-set). But a P-rational sequence of steps could not be defined as one each step of which is P-rational. For X may be P-rational given Y, Y P-rational given W, without X being P-rational given W, and it would be odd indeed, if we had to accept that a P-rational sequence of steps could lead from some intentional state to something which it doesn't rationally permit. For a sequence of steps to be P-rational, each of the terms (action or intentional states) involved must be P-rational given the set of all terms on which it causally depends, and this is ensured by the condition laid down above. Yet it would not be sufficient to say that a process is P-rational iff every sequence of steps it contains is P-rational. This is because several sequences of steps may (and will normally) converge on the same action/belief. To take the simplest case, suppose A and B are both immediate partial causes of X, then X may be P-rational given A and P-rational given B, without being P-rational given the set containing both A and B; in such a case, the two sequences of steps involved in the process will be P-rational, but one would not want to say that the process itself is P-rational. This is also taken care of by the condition stated above.

This is a good place to note that it is not required that the members of G (or G#) be P-rational given any subset of G (or G#). But, should it turn out that some member of G (or G#) is not P-rational given the set of all other members of G (or G#), the defect should not be seen in the process leading from G to X, but in the basis (or total basis) of this process. With these remarks in mind, let us now return to the explanation of agent rationality.

First note that proposal (7) above should now be read as:

(9) S is P-rational (at t) in doing/believing X iff (i) S's doing/believing X results from a P-rational process starting from a P-rational basis, (ii) X is P-rational for S given T-minus-X, and (iii) S is P-rational in being in T-minus-X.

But given the way in which the notion of process has been explained, almost any single action/belief turns out to be the outcome of many different processes starting from many different bases, and the condition that S's doing/believing X must result from *a* P-rational process starting from a P-rational basis now appears to be too weak. It has to be said exactly which of these processes and bases we are talking about and are relevant to agent rationality.

Let G be the largest subset of T on which S's doing/believing X is based, then *the* process leading from G to X is the maximal process from which S's doing/believing X results, in the sense that any process from which it may also result will be a sub-process of this process. Since G is the largest subset of T on which S's doing/believing X is based, it will be identical to the set G# of all intentional states of S involved in the (maximal) process leading from G to X and form the total basis of this process, while the set G* of members of G that are not based on any other intentional state of S will form its proper basis. The most plausible way to revise (9) will then be to say that[19]:

(10) S is P-rational (at t) in doing/believing X iff (i) S's doing/believing X results from a P-rational maximal process, the total and proper bases of which are P-rational for S, (ii) X is P-rational for S given T-minus-X, and (iii) S is P-rational in being in T-minus-X.

This proposal would obviously give rise to a regress if S's being P-rational in doing/believing X always depended on his being P-rational in doing/believing something

---

[19] It would be odd to claim that S can be P-rational in doing/believing X, even when his doing/believing X rests on a set of properly basic states that is not P-rational. And this condition has to be made explicit since, as will appear below, from the fact that the total basis of a process is P-rational, it doesn't follow that its proper basis is.

else. But suppose that T-minus-X is P-rational for S and contains only states that are not (at t) based on any other (though they might once have been based on mental states of S). This means that every member of T-minus-X is causally sustained by something other than a present intentional state of S, and I would submit that, in such a case, either it makes no sense to ask whether S's being in any of these states results from a rational process, or it must be granted that such ("zero-step") processes always count as P-rational. This would seem to be sufficient for S to be P-rational in being in each member of T-minus-X. Since at any given time, some of the agent's intentional states are not based on any other intentional state of his (and could thus be called "properly basic for S at t")[20], our current explanation of agent rationality will appear to be grounded, once it is admitted that all properly basic states result (by default) from a P-rational maximal process the total and proper bases of which are P-rational[21].

However, since to admit that properly basic states always result from P-rational maximal processes means that there are conditions under which S is P-rational in being in T-minus-X iff T-minus-X is P-rational for S, it might be feared that, given the way in which process rationality has been explained, it will turn out that this is *always* the case, and hence that condition (iii) in (10) says no more than that T-minus-X is P-rational for S. This would be the case if it turned out that

(11) if X is P-rational for S given T-minus-X then it is P-rational for S given
any subset of T-minus-X.

For suppose that T-minus-X is P-rational for S. Under assumption (11), this would imply that each member of T-minus-X is P-rational for S given the set of all other members of T-minus-X, *as well as* given any subset of this set. And this, in turn, would imply that every member of T-minus-X is the outcome of a P-rational maximal process starting from P-rational (total and proper) bases; for the bases of any such process can only be subsets of T-minus-X. It would be hard to see, in these conditions, how S could fail to be P-rational in being in T-minus-X.

But it can easily be seen that (11) is just as false as

---

20 I note in passing that this doesn't imply that it can be P-rational to do something for no reason, if actions must be caused by intentional states; but it does imply that it can be P-rational to believe something for no reason, if it is granted that some intentional states are not caused by any others (and hence are "psychologically" primitive or basic).
21 Provided only that their unit-set is P-rational for S.

(12) if X is P-rational for S given some subset of T-minus-X then it is P-rational

for S given T-minus-X,

and basically for the same reason. For no one will deny, for example, that (*ceteris paribus*) it may be P-rational for S to wear his sunglasses/believe that it is not going to snow this afternoon, given that he believes that it is sunny, while it would not be P-rational for him to wear his sunglasses/believe that it is not going to snow this afternoon, given that he believes both that it is sunny and that the radio said that it was going to snow. *By the same token*, it may be P-rational for S to refrain from wearing his sunglasses/believing that it is not going to snow, given that he believes both that it is sunny and that the radio said that it was going to snow, while it would not be P-rational for him to refrain from wearing his sunglasses/believing that it is not going to snow, given only that he believes that it is sunny. It follows that (11) and (12) are both false, that a set of intentional states can be P-rational even though not all its subsets are, and (as should have already been obvious) that a set of intentional states can be P-rational even though it is included in a larger set which is not. In other words, (action/belief) P-rationality is no more anti-monotonic (or "downward monotonic") than it is ("upward") monotonic[22].

This has the somewhat surprising and unnatural consequence that S can be P-rational in being in each member of a given set of intentional states without being P-rational in being in this set of states. For suppose that S is P-rational in being in T-minus-X, and that some subset G of T-minus-X is not P-rational for S. It has been laid down above that T-minus-X should then be P-rational for S, and it is natural to say that S is then P-rational in being in *each member* of T-minus-X. But then, S will also be P-rational in being in each member of G, and since *ex hypothesi* G is not P-rational for S, S cannot be P-rational in being in G! But if this is so, an explanation of when S is P-rational in doing/believing something will not suffice to provide an explanation of when S is P-rational in being in T-minus-X, and such an explanation is needed to ensure that condition (iii) in (10) will make sense. Perhaps this problem can be solved by laying it down that S is P-rational to be in some set G of intentional states iff S is P-rational to be in each member of G *and* G is P-rational for S. The upshot is that (10) can now be reformulated as:

---

22 It is therefore not governed by the same "logic" as consistency, which is downward, but not upward, monotonic. In other words, if a given set of beliefs is consistent, all its subsets are, but not all its supersets. I am much grateful to François Lepage for having helped me to realize this.

(13) S is P-rational (at t) in doing/believing X iff (i) S's doing/believing X results from a P-rational maximal process, the total and proper bases of which are P-rational for S, (ii) X is P-rational for S given T-minus-X, (iii) T-minus-X is P-rational for S, and (iv) S is P-rational in being in each member of T-minus-X.

But this explanation (even more clearly than (10)) involves a vicious circularity. For suppose that Y belongs to T-minus-X (at t). According to (13), and independently of whether there is any causal (or sustaining) link between X and Y (and hence, even if X and Y are both properly basic at t), whether S is P-rational (at t) to be in X depends on whether S is P-rational (at t) to be in Y. But this, in turn, may depend on whether S is P-rational (at t) to be in X, since X may (and will, when it's not an action) also belong to the set T-minus-Y of all other intentional states of S (at t)[23]. Condition (iv) in (13) must therefore be dropped, or at least replaced by a weaker one, such as the condition that the agent be P-rational in being in each member of the (total, and hence proper as well) basis of the maximal process from which his doing/believing X results; yielding:

(14) S is P-rational (at t) in doing/believing X iff (i) S's doing/believing X results from a P-rational maximal process the total and proper bases of which are P-rational for S, (ii) X is P-rational for S given T-minus-X, (iii) T-minus-X is P-rational for S, and (iv) S is P-rational in being in each member of the total basis G# of the maximal process from which his doing/believing X results.

In either case, it will have to be admitted that it was confused to suggest, as I did in the preceding section, that S could fail to be P-rational (at t) in doing/believing X, even though his doing/believing X resulted from a P-rational process starting from a P-rational

---

[23] It will obviously not do just to replace condition (iv) with the condition that, just before t, S *was* P-rational to be in T-minus-X. For this would mean that the set T*-minus-X, obtained by removing X from the set T* of all intentional states in which S was just before t *must* be identical to T-minus-X; and hence, that nobody could change his mind. Notice that the problem is not in allowing these two sets to be the same, but in *requiring* them to be the same. For suppose they just happen to be the same. Then (assuming that S was in Y at t-1) whether S is P-rational at t to be in X will depend (*inter alia*) on whether S was P-rational at t-1 to be in Y. But even if S already was in X at t-1, S's being P-rational at t-1 to be in Y may not depend on whether S was P-rational at t-2 to be in X, for S may well not have been in X at t-2, but only on whether he was P-rational at t-2 to be in the set of all intentional states in which he was then. No problem would arise even if S had always been in X, provided that he has not always been in Y. There would be a problem only if S has always been in both X and Y, but I think this can safely be ignored.

So one way to avoid the circularity would be to replace condition (iv) with the condition that, just before t, S was P-rational to be in T*-minus-X. But such a move would imply that an agent can be P-rational in doing/believing something only if he has always been P-rational in doing/believing anything that he ever did/believed. Furthermore, it is hard to see why S's being P-rational at t to be in X should depend on his having been P-rational to be in a certain state in

basis, X was P-rational for S given T-minus-X and T-minus-X was P-rational for S, i.e., just because he failed to be P-rational in being in each member of T-minus-X. What is left of this suggestion, under the current proposal, is (at most) that S must be P-rational to be in each member of G#. So it could happen that S is P-rational in doing/believing X, while he fails to be P-rational in being in some intentional state Z, not belonging to G#. But what would be wrong with that, since this means that S is not P-rational to be in some state on which his doing/believing X doesn't depend at all? Well, *perhaps* it could be claimed that what is wrong is that if S had been P-rational at some point in the past, he would not find himself in state Z, and this may have had the result that the set obtained by removing X from the set of all his present mental states would not have been P-rational. But this kind of claim doesn't make much sense, since it would seem no less likely that if S had been P-rational at some (other) point in the past, he could have been P-rational to be in state Z.

Intuitively, we want it to be the case that, if S is P-rational in doing/believing X, then the set T itself (i.e., the result of adding X to T-minus-X) is P-rational for S. But conditions (ii)-(iii) in the foregoing don't guarantee that, and leave open the possibility that, as a result of being P-rational in doing/believing something, the agent will find himself in a (new) total mental state which is not P-rational (even though he was in a P-rational total state "just before"; i.e., even though T-minus-X is P-rational) and this sounds odd. This is avoided by replacing conditions (ii)-(iii) in (14) with the condition that the set T of *all* intentional states (or actions) of S be P-rational for S. This would entail that X will be P-rational for S given T-minus-X (the former condition (ii)), as well as that each member Y of T-minus-X will be P-rational for S given T-minus-Y, and not only given (T-minus-X)-minus-Y.

Of course, this requires that it makes sense to say that an intentional state is rational/irrational given what the agent is doing (and not only given that he is in such and such intentional states). The idea sounds somewhat counterintuitive, but I am prepared to give it a try (at least until it becomes clear that and why actions require separate treatment)[24]. More importantly, the same point forces us to revise our explanation of

---

which he may well no longer be (especially since the envisaged move doesn't require X to be P-rational given T*-minus-X, but only given T-minus-X).

24 In any case, the idea seems less objectionable when it is recalled that we are here talking about intentional actions, and that many authors are explicitly committed to the (so-called "simple") view that S intentionally does X only if he intends to do it. This suggests that it would not be altogether meaningless to say that it is P-rational for S to intend to do X, given

process rationality; for just as we want it to be the case that if S is P-rational in doing/believing X then T is P-rational for S, we want it to be the case that if S's doing/believing X is *rationally* based on some set G of intentional states of his, then the set G#-plus-X (i.e. the set obtained by adding X to the total basis of the process leading from G to X) is P-rational for S. Accordingly, our earlier characterization of process rationality (i.e., (8)) must be replaced by the following:

(15) the process leading from G to X is P-rational for S iff (i) the set G#-plus-X is P-rational for S, and (ii) each member of the total basis G# of this process is such that the set obtained by adding it to the largest subset of G# on which it is based is P-rational for S.

Taking these changes into account, (14) can be reformulated as:

(16) S is P-rational (at t) in doing/believing X iff (i) S's doing/believing X results from a P-rational maximal process, the total and proper bases of which are P-rational for S, (ii) T is P-rational for S, and (iii) S is P-rational in being in each member of the total basis of the maximal process from which his doing/believing X results.

It can easily be seen that conditions (ii)-(iii) (in 16) do not entail condition (i), (i.e., that the fact that T is P-rational for S and S is P-rational in being in each member of G# doesn't entail that the total and proper bases of the maximal process from which his doing/believing X results are P-rational for him), which ensures that the notion of process rationality (and with it the distinction between agent and action/belief rationality) doesn't threaten to collapse. For suppose that T is P-rational for S and S is P-rational in being in each member of G# (the total basis of the maximal process from which X results). It follows from this that each member Y of G# results from a P-rational maximal process the total and proper bases of which are P-rational (and thus that the largest subset of G# on which Y is based is P-rational for S), but not that G#-plus-X, or the set G* of properly basic states of G#, or G# itself are P-rational for S (since, as was shown above, that S is P-

---

that he is (intentionally) doing it. It must also be kept in mind that the notion of rationality here in question is not that of agent rationality but that of action/belief rationality, and that it doesn't follow from what has just been suggested that someone's being in a certain intentional state can ever be rationally based (in part) on what he is (intentionally) doing.

rational in being in each member of a given set of states doesn't entail that this set is P-rational for S)[25].

3. Concluding worries

It may seem too strong to require that S can be P-rational in doing/believing X only if T is P-rational for him. But it must be remembered that this condition doesn't imply that the set T-minus-X has to be P-rational. Hence, this proposal is consistent with the intuition that if, for whatever reason, S should at some point find himself in a total mental state T which is not P-rational, then (there is a sense in which) it must still be possible for him to be P-rational in doing/believing something (though not something already belonging to T), but this could happen only if his so doing/believing results in his being in a (new) P-rational total mental state. It must also be kept in mind that agent rationality may often be confused with process rationality, and that (at least some of) the cases where we are tempted to say that the agent is P-rational in doing/believing X despite the fact that his total mental state is not P-rational, can be described as cases where either the maximal process leading to X or its (proper or total) basis, or both, are P-rational (and it could be claimed that in many contexts, this is what mostly matters).

A closely related worry may arise from the fact that, according to (16), for S to be P-rational in doing/believing X, it must be the case both that the total basis of the maximal process leading to his doing/believing X is P-rational, and that the set obtained by adding X to this total basis is P-rational. But given that a set of intentional states is P-rational if and only if each of them is P-rational given all the others (or more strictly speaking, given the

---

25 It may be worth trying to illustrate this point by describing a "concrete" case. Let's suppose that $\{a1…an, b1…bm, c1…cp, X\}$ is the set T of all intentional states of S at time t, where a1…an (and only them) are properly basic states of S (at t), that the total basis G# of the maximal process on which X is based is $\{a1, a2, a3, b1, b2, c1, c2\}$, and hence that the proper basis G* of this process is $\{a1, a2, a3\}$. Suppose further that a1 is a partial cause of b1, b2, and c1, a2 is a partial cause of b2, a3 a partial cause of c2 and b1 a partial cause of c1. Then, according to (16), to say that S is P-rational (at t) in doing/believing X is to say that:

T is P-rational for S (i.e. that each member of T is P-rational given all other members of T)

G# is P-rational for S

G* is P-rational for S

G#-plus-X is P-rational for S

$\{a1, b1, c1\}, \{a1, b1\}, \{a1, a2, b2\}, \{a3, c2\}, \{a1\}, \{a2\}, \{a3\}$ are P-rational for S

S is P-rational to be in each member of G#, which means that each of them must result from a P-rational maximal process starting from a P-rational total basis, each member of which S is P-rational to be in (in terms of our illustration, this boils down to the further requirement that G#c1 = $\{a1, b1\}$ and G#b2 = $\{a1, a2\}$ are P-rational for S).

Now, to see that conditions (ii)-(iii) in (16) do not entail condition (i), it suffices to see that all of T, $\{a1, b1, c1\}, \{a1, b1\}, \{a1, a2, b2\}, \{a3, c2\}, \{a1\}, \{a2\}, \{a3\}, \{a1, b1\}$ and $\{a1, a2\}$ can be P-rational while G*, G# and G#-plus-X are not.

set obtained by removing it from the given set), and assuming that an intentional state is P-rational given some set G of intentional states if and only if no better alternative is left open by G, one may well wonder how an agent could ever be rational to add anything to a *rational* set of intentional states. To see more clearly what I have in mind, note that there is nothing to prevent the total basis of the maximal process leading to S's doing/believing X to be identical with T-minus-X. When this is so, for S to be P-rational in doing/believing X, it has to be the case that both T and T-minus-X are P-rational for S. Obviously, this can be the case only when X is not an alternative to any member of T-minus-X (for if it were, it could not be the case that T-minus-X is P-rational). But as far as I can see, this may very well happen; which ensures that an agent whose total mental state already is P-rational can still be P-rational to do/believe something else[26].

This brings us back to our informal explanation of action/belief rationality as "being at least as good as any other alternative". Putting to one side the question of what could be meant here by saying that some alternative is good, or better than another, it is obviously crucial to have a conception of what makes it the case that some action/belief types are alternatives to each other. And here, a number of questions arise, which I unfortunately cannot discuss in detail. It has to be remarked, however, that the conditions for agent rationality which have been put forward above don't rest specifically on this understanding of action/belief rationality, and that there may possibly be other ways of explaining the notion of "rational permission" on which they rely which would not make them less acceptable.

Insofar as action/belief rationality is supposed to be (no doubt partially) explained by saying that an action/belief type is P-rational (given some set T of intentional states) iff it is at least as good as any other alternative, it would be circular to say that action/belief types X and Y are alternative to each other iff (i) S could do/believe X, (ii) S could do/believe Y but (iii) it would not be P-rational for S to do/believe both X and Y. Taking this line would then force us to take the notion of P-rationality as primitive, with no guarantee that (when the notion of alternative is defined as just suggested) it will turn out

---

[26] Though it remains to be seen whether this is enough to account for all the cases in which we would like to say that an agent whose total mental state already was rational has been rational to change his mind.

(or even that it would make sense to say) that an action/belief type is P-rational iff it is at least as good as any other alternative[27].

It must therefore be possible to say when two action/belief types count as alternative to each other without presupposing the concept of rationality. Let us think first of actions. It seems at once plausible and natural to hold that two action types are alternative to each other (for S) iff S can do each of them, but not both. It is worth pointing out that an apparent advantage of this conception is that it doesn't prevent two alternative actions from being nonetheless jointly P-rational. For suppose S wants both to buy an ice cream and to buy an apple, but that he has enough money to buy only one of these. Suppose further that, given his beliefs and preferences, buying an apple is better than buying an ice cream. Then it would be P-rational for S (given his intentional states) to buy an apple, and rationally prohibited (i.e., not P-rational) for S to buy an ice cream. But it clearly doesn't follow from this that it would not be P-rational for S to buy both an apple and an ice cream, if he had more money or if prices were lower. So what makes buying an apple and buying an ice cream alternative actions is just the fact (in this case a contingent fact) that S doesn't have the capacity to do both, and *cannot* be the fact that it would not be P-rational for S to do both (for this is not a fact).

The trouble is that it is not obvious how to generalize this conception to beliefs and other intentional states. For one thing, there seems to be far less constraints on what an agent can simultaneously believe or desire than on what he can simultaneously do. For another, if there are such constraints, they are likely to be largely specific to each agent, while there are fairly general limitations on what any normal agent can simultaneously do. Maybe it could be suggested that a normal agent is unable to have logically incompatible beliefs (and thus that if belief types X and Y are logically incompatible, then they are alternative to each other). But this would mean that no belief system (of a normal agent) could fail to be P-rational in virtue of containing logically incompatible beliefs; and this sounds odd, in view of the fact that logical inconsistency is often thought of as the (or a) paradigmatic form of irrationality.

---

[27] This point is unaffected by the observation that "to be as good as any other alternative" might turn out to be nothing but "to be as rational as any other alternative". For this comparative notion of rationality is not the notion of P-rationality here in question; and as far as I can see it may well be the case that the notion of P-rationality rests on some comparative notion of rationality. The point is that this could not happen if the notion of alternative is defined in terms of P-rationality.

Suppose that, at any given time, S must be executing one and only one action in each of the following sets of "alternatives": {A1, A2, A3}, {B1, B2, B3}, {C1, C2, C3}. This means that at any given time, S is doing either (A1, B1, C1), or (A1, B1, C2), or (A1, B1, C3), etc... Now suppose that S is doing (A1, B1, C1) at t, and it is asked whether it is P-rational for S to do A1 at t, given T-minus-A1. In terms of our suggested explanation of action/belief P-rationality, this amounts to asking whether doing A1 is at least as good for S as doing A2 or doing A3, given all information about what else S is doing/thinking or not doing/thinking (at t). This serves to highlight the fact that the alternative to doing something doesn't normally consist in just not doing it (or refraining to do it), but in doing any one of a certain number of other things (that are excluded by his doing it). Notice that when removing the fact that S is doing A1 from T, one is also removing the fact that S is not doing A2 or A3 (which means that T isn't really a set of states or actions, but a complete specification of what the agent does/think and of what he doesn't do/think, but such a specification is implicitly determined by the set of all intentional states/actions of the agent), since *ex hypothesi*, S's doing A1 entails his not doing A2 and his not doing A3; so what really gets evaluated is (A1, not-A2, and not A3); but one is not removing the fact that S is not doing B2 or B3 instead of B1.

Can something similar be said in the case of beliefs? Does the fact that S believes p entails that he believes or that doesn't believe any other specific thing? I.e., when removing the fact that S believes p from T, is one perforce removing the fact that S believes some other thing, or the fact that he doesn't believe some other thing? It is a brute physical fact that no agent can, say, move one of his fingers in two directions at once. But are there similarly brute psychological facts such as that no agent can believe both p and q, for some p and q? (of course, for any given agent, it might be that there are some things that he is, as a matter of brute psychological fact about him, simply unable to believe at once, but the same holds for actions, a given agent may be unable to simultaneously do certain things that other agents can) But there is no guarantee that for every agent, and everything he might believe, his believing it entails that he doesn't believe certain other things. But then nothing excludes the possibility that for some agent who believes p, the only alternative to his believing p, will be not to believe p. And this is not an intentional state, though it is an intentional fact about the agent (it is a fact concerning his being or not being in a certain

intentional state). This possibility isn't specific to beliefs or intentional states, since it may well happen that, for a given agent and a given action, the only real alternative to his doing it is his not doing it (e.g. suppose there is just one way in which S can move a given finger; then the only alternative to his moving it is his not moving it at all).

But then, unless we are prepared to claim that not doing something is an action and not believing something is an intentional state, we have to withdraw the suggestion that only actions and intentional states (beside agents) can be said to be rational/irrational. It could still be maintained, however, that only things that can be said to be rational/irrational can make other things rational/irrational, and that only facts concerning what an agent does or thinks or doesn't do or doesn't think can be rational/irrational. In other words, it could be maintained that the fact that S doesn't do/believe something is an intentional fact about S, and that as such, it can both be said to be rational/irrational and contribute to make other things rational/irrational.

References

Alston, William P. (1985) "Concepts of Epistemic Justification", Alston (1989) 81-114
Alston, William P. (1989) *Epistemic Justification*, Ithaca, Cornell U. Press
Audi, Robert (1985) "Rationalization and Rationality", Audi (1993b) 405-430
Audi, Robert (1993a) *Action, Intention, and Reason*, Ithaca, Cornell U. Press
Audi, Robert (1993b) *The Structure of Justification*, Cambridge, Cambridge U. Press
Bratman, Robert (1987) *Intention, Plans and Practical Reason*, Cambridge (Mass.), Harvard U. Press
Firth, Roderick (1978) "Are Epistemic Concepts Reducible to Ethical Concepts?", Firth (1998) 237-249
Firth, Roderick (1981) "Epistemic Merit, Intrinsic and Instrumental", Firth (1998) 259-271
Firth, Roderick (1998) *In Defense of Radical Empiricism*, Lanham, Rowman & Littlefield
Foley, Richard (1987) *The Theory of Epistemic Rationality*, Cambridge (Mass.), Harvard U. Press
Foley, Richard (1993) *Working without a Net*, Oxford, Oxford U. Press
Hubin, D. Clayton (1980) "Prudential Reasons", *Can. J. of Phil.* 10, 63-81
Hurley, Susan L. (1989) *Natural Reasons*, Oxford, Oxford U. Press
Goldman, Alvin I. (1988) "Strong and Weak Justification", Goldman (1992) 127-142
Goldman, Alvin I. (1992) *Liaisons*, Cambridge (Mass.), MIT Press
Kvanvig, Jonathan L. and Christopher Menzel (1990) "The Basic Notion of Justification", *Phil. Studies* 59, 235-261
Pettit, Philip and Michael Smith (1993) "Practical Unreason", *Mind* 102, 53-79