OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Mind Association

Minimal Rationality
Author(s): Christopher Cherniak
Source: *Mind*, New Series, Vol. 90, No. 358 (Apr., 1981), pp. 161-183
Published by: Oxford University Press on behalf of the Mind Association
Stable URL: http://www.jstor.org/stable/2253336
Accessed: 10/03/2009 20:28

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=oup.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

Oxford University Press and *Mind Association* are collaborating with JSTOR to digitize, preserve and extend access to *Mind.*

http://www.jstor.org

Mind (1981) Vol. XC, 161–183

Minimal Rationality[1]

CHRISTOPHER CHERNIAK

In intentional explanations of behaviour, we require rationality of the agent. How rational must a creature be to be an agent, that is, to qualify as having a cognitive system of beliefs, desires, perceptions? In the philosophy of psychology, there has been a relatively uncritical acceptance of highly idealized conceptions of rationality. I shall attempt here to characterize a concept of *minimal* rationality; in particular, according to such a rationality concept, an agent can have a less than perfect deductive ability. I shall argue that we in fact require only minimal, as distinguished from ideal, rationality of an agent. I shall further propose that such minimal rationality conditions are indispensable for any adequate intentional theory. What is at stake is closely related to issues concerning the very possibility of a cognitive science, and of a realist interpretation of it.

1. *The 'Autonomy of the Mental'*

In 'A Scandal in Bohemia', Sherlock Holmes' opponent has hidden a very important photograph in a room, and Holmes wants to find out where it is. Holmes has Watson throw a smoke bomb into the room and yell 'fire' when Holmes' opponent is in the next room, while Holmes watches. Then, as one would expect, the opponent runs into the room and takes the photograph from where it was hidden. Not everyone would have devised such an ingenious plan for manipulating an opponent's behaviour; but once the conditions are described, it seems very easy to predict the opponent's actions. *Prima facie*, we seem to predict the actions not as common-sense behaviourists or neurophysiologists, but by assuming that the opponent possesses a large set of beliefs and desires—including the desire to preserve the photograph, the beliefs that fire will destroy it, that where there's smoke there's fire, etc.—and that the opponent will act appropriately for those beliefs and desires. It seems an uncontroversial fact that we very commonly employ this procedure for predicting people's behaviour in everyday situations. A less exotic example than the Holmes' story is that often when I step

1 I am grateful to Charles Chihara, William Craig, Daniel Dennett and Barry Stroud for their generous assistance.

into a crosswalk, I am betting my life (not always with complete equanimity) on expectations of a motorist's behaviour that seem to be based on assumptions regarding his perceptions, beliefs, and desires.

The idea that there can be a predictive theory of belief is of course not unfamiliar. However, a long and central tradition in the philosophy of mind denies the possibility, even in principle, of such predictions of behaviour on the basis of attribution of a cognitive system. Descartes' distinction between the domains of the physical and the mental, where the former is subject to deterministic laws while the latter is free and not subject to any laws, has such a consequence. This type of view continues to be significant in the current debate concerning the possibility and nature of cognitive psychology. I think the influence of the view can be perceived in D. C. Dennett's important paper 'Intentional Systems'.[1] At the beginning of the paper, Dennett proposes to examine the concept of 'a system whose behaviour can be (at least sometimes) explained and predicted by relying on ascriptions to the system of beliefs and desires' (p. 87). And yet there is a tension present; by the end of the paper, Dennett is claiming, 'If one wants to predict and explain the "actual, empirical" behaviour of believers, one must . . . cease talking of belief and descend to the design stance or physical stance for one's account.'[2] Donald Davidson has also defended a quasi-autonomist position by arguing that there can be 'strict' or 'serious' deterministic laws only in the physical domain, and that intentional theories cannot yield 'accurate' predictions.[3]

There is some conflict between the simple fact of the success of common-sense psychological practice and philosophical assertions of the autonomy of the mental, even when that autonomy is attenuated to a matter of degree of precision of laws. The purpose of this paper is to suggest one source of this conflict. I shall not attempt to disprove every one of the many arguments for the impossibility of a predictive intentional theory; I shall show only that one crucial precondition for an intentional theory with pre-

1 *Journal of Philosophy* (1971).

2 In a draft of a recent unpublished paper 'Three Kinds of Intentional Psychology', Dennett seems to have rejected this latter view more unequivocally; the commitment to ideal rationality remains.

3 See 'Actions, Reasons, and Causes', *Journal of Philosophy* (1963); 'Psychology as Philosophy', in J. Glover (ed.), *The Philosophy of Mind* (Oxford, 1976); 'Mental Events', in L. Foster and J. W. Swanson (eds.), *Experience and Theory* (London, 1974).

dictive content has been mistakenly rejected by some of the most influential of these positions. The main hypothesis here is that the pervasively and tacitly assumed conception of rationality in philosophy is so idealized that it cannot apply in an interesting way to actual human beings. One of the most significant consequences of such an extreme idealization is that it tends to exclude a realist account of mental entities in favour of an instrumentalist account (such as Dennett's): if the only possible rationality conditions on, e.g., beliefs are so idealized as to be inapplicable to humans, then any attributions of beliefs to humans cannot really be *true*; the attributed entities are at most useful myths. I shall be exploring the implications of the concept of minimal rationality, where the agent has a less than perfect ability to choose appropriate actions. I shall be principally concerned with rationality conditions on belief sets, and on the believer's deductive logical abilities. The strategy in approaching the question of what minimal rationality is will be to proceed by successive approximations.

2. *Two types of belief theory*

Rationality conditions can be either too weak or too strong for a satisfactory predictive theory. Before we consider theories of belief that presuppose ideal rationality, let us examine the opposite type of theory, one that requires no rationality at all of a believer. The most rudimentary theory of this kind is an *assent theory of belief*:

> An agent believes all and only those statements which he would affirm.

Russell's theory of belief in 'On Propositions'[1] included an internalized version of this public assent criterion of belief. According to Russell, a person's believing a proposition at a particular time consists of the believed proposition occurring with a 'feeling of assent' to the proposition in the person's psychological history at the time.

Such a theory has an attractive simplicity and perhaps tends to satisfy feelings that a believer is the final authority on what his beliefs are. However, a crucial defect of the assent theory as a complete belief theory is that it does not impose any rationality constraints upon the belief set, or more than vestigial constraints

1 In R. Marsh (ed.), *Logic and Knowledge* (New York, 1971); see p. 311. A similar theory is proposed in Bruce Vermazen's 'Consistency and Underdetermination', *Philosophy and Phenomenological Research* (1968).

upon the relation of the belief-desire set to the agent's actions. It is a 'null' rationality requirement. According to the assent theory, no inferences, however obvious and useful, need be made from the beliefs, and the belief set can include any and all inconsistencies. The belief-desire set is not required to guide at all the choice of appropriate actions, except for the small area of verbal behaviour of assent and dissent. Consequently, the assent theory does entail the autonomy of the mental domain, and makes a mystery of our everyday successes in predicting behaviour on the basis of belief-desire attributions. A belief theory with no rationality restrictions is without predictive content; using it, we can have virtually no expectations regarding a believer's behaviour.

In contrast to the permissiveness of the assent theory, the prevalent accounts of belief have included conditions requiring ideal rationality of an agent. In decision and game theory, the principle that the agent will generally choose (or, is likely to choose) the action which maximizes his expected utility is commonly recognized as being of this type. While this principle does have valuable applicability for certain ranges of problems, there has been some realization in decision and game theory relatively recently of the serious difficulties that models assuming perfect information or rationality of an agent have as idealizations.[1] But in philosophy of mind and action and in theory of knowledge this issue has received little attention. The philosophical accounts employ, usually tacitly, an *ideal general rationality condition*, which we can formulate roughly as:

> If an agent has a particular belief-desire set, he would undertake *all* actions which are apparently appropriate.

As a convenient simplification here, we shall say that an action is *apparently appropriate* if and only if, according to the person's beliefs, it would tend to satisfy his desires. (A weaker ideal rationality condition is: an agent would undertake some nonempty set of apparently *most* appropriate actions.)

Such an idealized theory of belief may be of some value as a convenient simplification of the theory we actually employ in everyday situations, but otherwise it is unacceptably stringent. For, this ideal rationality condition is generally unrealizable. Of course, it would exclude the possibility of someone having beliefs and

1 One of the earlier, and most influential, discussions is in Part IV of H. Simon's *Models of Man* (New York, 1957).

desires and even occasionally being forgetful or careless in his choice of actions. Consequently, with such a theory, Holmes could not have predicted his inevitably forgetful and careless opponent's behaviour on the basis of an attribution of a belief-desire set; he would have had to regard the opponent as not having any cognitive system. But as we shall discuss later, this ideal rationality condition requires a believer not only not to be sloppy, but to have a peculiarly idealized deductive ability.

The unsatisfactoriness of the ideal general rationality condition arises from its denial of a fundamental feature of human existence, that humans are in the *finitary predicament* of having a fixed limit on their cognitive capacities and the time available to them. Unlike Turing machines, actual humans in everyday situations or even in scientific inquiry do not have potentially infinite memory and computing time. Since any human is in the finitary predicament, using a belief theory with this rationality condition amounts to having no applicable intentional theory at all. The basic limitations imposed by the finitary predicament are not confined to creatures with our particular intellectual abilities. The limitations are general in that they would be just as unavoidable, for example, for a creature that had available the resources of the entire galaxy and all of the time until heat-death of the universe. Thus, any assumption that the ideal general rationality condition is harmless, in that human rationality approximates it, must be closely examined.

3. Minimal rationality

The value of an idealization is always relative to a set of goals. The ideal general rationality condition is useful under some conditions, as Von Neumann and Morgenstern, for instance, explained for their game-theoretic models in *Theory of Games and Economic Behavior*.[1] A sound motivation for idealizing a theory is that the resulting simplification yields a theory which is more manageable (e.g., for the purposes of formalization) than an entirely correct or complete theory would be. A theory can be idealized to different degrees. Consequently, simplicity and manageability can be traded off for greater applicability. The theory of belief based upon minimal rationality conditions which I propose is an attempt to obtain some significant applicability in exchange for a more complex theory. The theory of belief here continues to be significantly

1 Princeton, 1944.

idealized in the conception of inference. The account will be principally concerned with verbally formulated beliefs. I will generally be treating an agent's beliefs as a set of sentences, and an inference from those beliefs as the addition of a sentence to that set. I will also not distinguish between deliberate, conscious inferences and unconscious inferences. The rationality conditions below are only necessary conditions for having beliefs and desires. All of the other minimal rationality conditions presented below can be derived from a *minimal general rationality condition*:

> If an agent has a particular belief-desire set, he would attempt some, but not necessarily all, of those actions which are apparently appropriate.

The argument that, as belief-attributors, we in fact employ—and in addition, *should* employ—this minimal rationality condition is by exhaustion of a trichotomy. We have seen that we are able to predict people's behaviour on the basis of attributions of beliefs and desires, and that both an ideal and a 'null' rationality condition exclude this. For, we know that no creature in the finitary predicament can satisfy the ideal rationality condition. And if the believer were not required to be at least more likely to undertake some of the apparently appropriate actions, then the attribution of a belief-desire set could never yield any predictions of behaviour, and would never be disconfirmable by observed behaviour. On the basis of such an attribution, no behaviour could be expected; every action would be equally probable. In fact, recognition that some rationality condition on beliefs is required, combined with failure to distinguish minimal rationality from ideal rationality, gives the ideal rationality condition much of its 'all or nothing' plausibility. The only remaining possibility is a minimal rationality condition, so that is the rationality condition we must be using, and should use.

In addition, for a predictive belief theory, there is a stronger general rationality condition on a person's belief-desire set. The person must not only attempt some of the actions which are appropriate given his belief-desire set, but he must also *not* attempt enough of the actions which are *in*appropriate given that belief-desire set (consequently, a similar requirement accompanies the minimal inference condition described below). For, a creature whose behaviour is determined randomly is very likely, given enough time, to qualify as having any arbitrary belief-desire set,

according to the minimal general rationality condition without this additional requirement of 'negative rationality'. A related point is that the minimal general rationality condition is stronger than a purely extensional requirement, in that it has counterfactual implications.

The minimal general rationality condition implies that a believer must have a minimal deductive ability. (We shall not discuss here the question of whether satisfaction of the minimal general rationality condition requires inductive ability.) The *minimal inference condition* on deductive ability is:

> If an agent has a particular belief-desire set, he would make some but not necessarily all of the sound inferences from the belief set which are apparently appropriate.

These inferences need not involve verbally represented beliefs. For our purposes, the believer's undertaking an action appropriate for his beliefs and desires, where the beliefs and desires cause the action 'in the right way',[1] constitutes his concluding that the action is desirable; an entirely nonlinguistic creature like a young child can perform such inferences.

The minimal inference condition requires a believer to make some of the sound inferences which are apparently useful in selecting appropriate actions. If a believer did not satisfy at least the minimal inference condition, in that he would make no apparently appropriate inferences from his beliefs, he would not in general be able to recognize and undertake actions which were appropriate given those beliefs. For example, suppose the agent's putative belief set included the beliefs 'If it rains, then the dam will break' and 'It is raining'. The agent would never conclude that the dam will break, even if this would be obviously useful—for instance, when the person also believed he was below the dam and would be drowned if it broke, was not suicidal, etc. Therefore, the person would not be able to undertake any appropriate action on the basis of his beliefs (as opposed to, e.g., by whim) which depended on this information, such as fleeing. The person's deficit of logical insight, and consequently of rational action, would exclude him from having beliefs according to the minimal general rationality condition.

We can in turn describe the logical ability required by the

1 See Davidson's discussion of the distinctiveness of this type of causal efficacy in 'Actions, Reasons, and Causes'.

minimal inference condition in terms of a minimal appropriateness requirement on heuristic ability and a minimal consequence requirement on deducing ability. The *minimal appropriateness requirement* on which inferences the person would attempt to make is:

> The agent would undertake some of the sound inferences from his belief set which would be apparently appropriate for him to make.

That is, the resulting inferences would, according to his own beliefs, aid him in choosing other actions which would tend to satisfy his desires. (Objective appropriateness is not satisfactory here, since it would clearly be an unacceptably extreme idealization to regard the agent's beliefs as always correct.) The person must act as if he had made judgements of the form, 'According to my beliefs and desires, it would be useful for me now to know whether or not q is a consequence of my belief p', where some of these judgements are correct. Thus, a heuristic imbecile—for example, who just tries to deduce from a sentence p vacuous conjunctions $p \,\&\, p$, $(p \,\&\, p) \,\&\, p$, $((p \,\&\, p) \,\&\, p) \,\&\, p$, etc.—cannot have beliefs. The *minimal consequence requirement* on deducing ability is:

> The agent must succeed in performing some of the apparently appropriate sound inferences he has undertaken.

There is no implication, in stating separate appropriateness and consequence requirements, that two genuinely distinct processes always have occurred when an inference is made. The appropriateness and consequence requirements are interdependent, in that deducing ability (and also inductive ability) is required in selecting appropriate deductive tasks, and heuristic ability may be required (e.g., to identify useful lemmas) in order to perform complex deductive tasks already undertaken. The actual heuristic process of selecting apparently useful deductive questions is usually nonconscious. And there is a point of diminishing returns, beyond which there are other better uses of a person's time than in perfecting the choice of deductive questions to consider—for example, performing those alternative actions which will only be beneficial if performed at that time.

When we regard inferences as one variety of actions, the minimal appropriateness requirement is a special case of the minimal general rationality condition on actions. However, the selection of apparently useful inferences cannot itself be solely by means of

actual practical inferences—conscious or nonconscious—or there will be a regress.[1] As a first step toward avoiding this problem, we can say that in many cases, the person does not actually decide to undertake the inference. Rather, the conformance of actions of inferring with desires instead must arise largely by means of nonconscious mechanisms of selection or guidance that do not involve reasoning processes of any kind. These mechanisms may be acquired—for instance, as learned 'cognitive styles'—or the agent may be 'designed' by natural selection so that, as an efficient organism, he undertakes particular inferences.

4. Ideal deductive ability

The minimal inference condition is only a necessary condition for having beliefs. We shall not attempt to settle whether the minimal inference condition (with the additional 'negative rationality' condition described above) is a sufficient condition for being logically competent to have beliefs, that is, whether satisfaction of the augmented minimal inference condition constitutes possession of *all* of the logical ability required for having beliefs. However, we shall argue against conditions that require a believer to have ideal deductive ability. We first examine what ideal deductive ability is supposed to be, since there is a family of distinct types of 'ideal'.

The simplest, and most extreme, idealization is that an agent's belief set is *deductively closed*:

> An agent actually believes (or, infers, or can infer) all consequences of his beliefs.

This is the rationality idealization adopted in classical epistemic logic, notably in J. Hintikka's *Knowledge and Belief*.[2] But it is impossible for a person to infer and believe every one of the consequences of a belief, because there are too many, however they are individuated. This set of consequences includes the infinite set of all valid sentences expressible in the believer's language. In addition, most of these consequences are such that it is impossible to believe a single one of them; this is because each is so complex that it could not be stated in a lifetime, much less understood. Of course, any theory of belief that includes the deductive closure condition cannot apply to humans, or any other creature in the

1 Regresses of a similar type were discussed by Gilbert Ryle in ch. 2 of *The Concept of Mind* (London, 1949).

2 Ithaca, 1962.

finitary predicament. Hintikka himself explains that his axiomatization is applicable to the actual world 'only in so far as our world approximates one of the "most knowledgeable of possible worlds",' 'in which everybody follows the consequences of what he knows as far as they lead him' (p. 36)—that is, not at all.

In fact, the deductive closure condition requires more deductive ability than is needed to satisfy even the ideal general rationality condition (or related conditions, such as the principles of maximization of utility) introduced earlier. For, the deductive closure condition requires a believer to infer all consequences of his beliefs, whether apparently useful for him or not. The logical ability required for a creature to satisfy the ideal general rationality condition that a person would undertake all (not just some) actions appropriate for his beliefs is described by an *ideal inference condition*:

> If an agent has a particular belief-desire set, he would make all of the sound inferences from the belief set which are apparently appropriate.

This ideal inference condition is, for our purposes, equivalent to the following *ideal appropriateness* and *consequence requirements*, respectively, which correspond to the minimal appropriateness and consequence requirements above:

(i) An agent would select *all* those inferences to make from his beliefs that are apparently appropriate for him to make.
(ii) An agent would successfully perform *all* of those inferences.

If a creature did not satisfy both of these conditions, there might be actions which were apparently appropriate, but which the creature could not recognize to be appropriate, because it lacked the logical ability needed to identify them. For instance, the creature might (correctly) think its survival depended on its determining whether or not a particular sentence was a consequence of its beliefs, when in fact the creature could not perform this deductive task at all. We shall see below that Hintikka is not alone in accepting conditions requiring ideal deductive ability of a believer; proponents of the thesis of the autonomy of the mental generally tend to employ such idealizations. Acceptance of these idealizations excludes a predictive intentional theory.

It is important to see that, although the ideal inference condition is weaker than the deductive closure condition, it is still much too strong. The first problem is similar to one mentioned for the ideal general rationality condition earlier; it is a feature of our actual everyday belief-attributing practice that we do not deny that a person has a particular set of beliefs because he fails to infer from that set all apparently appropriate logical consequences—or even a feasibly small set of the apparently *most* useful consequences. Humans often fail to identify inferences of this kind; it's common to say, 'If I'd only *asked* myself whether q was true, I could have figured that out and then done . . .'. And people often cannot perform such deductive tasks even when they have identified them. As an example (from second-order logic), many have wanted all their lives to know whether Goldbach's conjecture is a consequence of accepted axioms of number theory, but this task has not yet been accomplished; nonetheless, we do not deny that these people accept the axioms.

Thus, we do not in fact use the ideal appropriateness and consequence conditions when we attribute beliefs to people in actual circumstances. Furthermore, the fact that we do not use these ideal conditions is not just an accident of our particular culture, like our lacking a one-letter word for 'all'; the choice between minimal and ideal conditions is not arbitrary. For, attributions of belief are not valueless for predicting behaviour if only minimal, rather than ideal, rationality is required. But adoption of the ideal conditions would prevent us from taking advantage of most of the opportunities for effectively predicting human actions on the basis of an attributed cognitive system, since we would then have *no* intentional theory which was actually applicable to humans. It is unfeasible, if not impossible 'in principle', for us to predict human behaviour to any significant extent on a purely neurophysiological or behaviouristic basis; only a cognitive theory can have the required level of abstractness. Hence, acceptance of the ideal conditions is a refusal to attempt to predict in almost all situations, a sulk because of human finitude.

Perhaps the clearest indication of the extreme inapplicability of the ideal conditions is that if a creature did non-vacuously satisfy them, most of the tasks of deductive sciences such as mathematics would be *trivial* for it. Failure to satisfy the ideal conditions thus should not be belittled as just a result of carelessness or sloppiness. The common comparison in economic theory of the principle of

maximization of utility with the postulation of dimensionless, perfectly elastic spheres by the ideal gas laws seems correspondingly inappropriate here. There is not even a significant probability above chance of the agent choosing many of the actions which would maximize his expected utility (e.g., when selecting the action is a deductive task of the difficulty of the Goldbach's conjecture case). And there is no better reason to expect deductive omniscience of a large population of agents—e.g., a scientific community—even if inquiry is pursued indefinitely, to some Peircean limit. In many situations, the relation of the ideal rationality conditions to actual humans is better compared not with the relation of the ideal gas laws to actual gases, but with the relation of phlogiston theory to actual gases. These departures of actual human behaviour from the idealization are less noticeable to the theorist, because he also cannot identify the appropriate actions. Those who supposedly use the idealization are thereby in effect often employing a theory of feasible inferences, as explained below.

The second way, then, in which the ideal inference condition is too strong, is that it excludes humans from having beliefs, and so adoption of it prevents virtually any prediction of human behaviour. A third way in which the ideal condition is too strong is that, for a non-suicidal creature in the finitary predicament, it would be irrational even to try to satisfy it. Generally, there would be much more desirable ways for him to use his limited cognitive resources (e.g., relating to immediate survival) than trying to ensure that every one of his actions is appropriate. We shall return to this point later.

5. *Minimal consistency*

The deductive ability required to satisfy the general minimal rationality condition must include not only an ability to perform useful inferences, but also an ability to eliminate inconsistencies in the belief set. The belief set is subject to a *minimal consistency condition*:

> If an agent has a particular belief-desire set, then if some (but not necessarily all) inconsistencies arose in his belief set, he would eliminate them.

This condition applies both to explicit inconsistencies such as $\{p, -p\}$, and to tacit ones such as $\{p, p \rightarrow q, -q\}$.

Like all of the earlier minimal rationality conditions, the minimal consistency condition is specified by exhaustion of a trichotomy: a believer cannot permit all inconsistencies in his belief set, but he should not be required to eliminate every inconsistency which might arise in his belief set; hence he must maintain minimal consistency. On the one hand, if an agent's cognitive system was not subject to some consistency constraint, and so could contain an unlimited number of inconsistencies, the attribution of such a system could not be of any value in predicting the agent's behaviour. We could never expect such an agent, in accordance with the general minimal rationality condition, to attempt an action appropriate for a given belief; for, briefly, this agent might in *any* case have another belief which was inconsistent with the given belief, and which he might then act upon instead. An intentional theory with constraints on only contradictions, and not on tacit inconsistencies, would still be without empirical content for the same reason.

On the other hand, the minimal consistency condition must be clearly distinguished from an *ideal consistency condition*:

> If an agent has a particular belief-desire set, then if any inconsistency arose in his belief set, he would eliminate it.

This consistency condition is unacceptable for the same three kinds of reasons as the ideal inference condition was. First, it is clear that we do not in fact employ such a condition in our intentional attributions; a person's having beliefs is not ruled out by the occurrence of a single inconsistency in his putative belief set. Second, adoption of the ideal consistency condition would not be advisable for the *attributor*, since it would amount to a refusal to attempt to predict behaviour in terms of a cognitive system for creatures of anything like the human level of logical abilities. For, this ideal condition restricts the class of believers not employing extremely conservative strategies of belief-acquisition to creatures for whom a large range of the tasks of the deductive sciences would be trivial. Third, the ideal condition requires a believer with our abilities and normal non-suicidal desires to be irrational, in that there are often epistemically more desirable activities for him than maintaining perfect consistency.

The ideal consistency condition gains plausibility from the recognition that some consistency constraint on beliefs is required, combined with a failure to distinguish minimal consistency from

ideal consistency. For instance, on the one hand, in 'Psychology as Philosophy' Davidson states, 'if we are intelligently to attribute attitudes and beliefs, or usefully to describe motions as behaviour, then we are committed to finding in the pattern of behaviour, belief, and desire a large degree of rationality and consistency'.[1] But on the other hand, the rest of Davidson's discussion strongly suggests that he thinks that the possession of beliefs and desires requires ideal consistency, rather than just 'a large degree' of consistency. For example, Davidson says, 'I do not think we can clearly say what should convince us that a man at a given time (or without a change of mind) preferred *a* to *b*, *b* to *c*, and *c* to *a*. The reason for our difficulty is that we cannot make good sense of an attribution of preference except against a background of coherent attitudes' (pp. 49–50). Davidson is assuming a particular 'ideal consistency condition' on preference—that transitivity is never violated.

Quine's Principle of Charity in the interpretation of a speaker's utterances—for instance, 'fair translation preserves logical laws'—seems historically to have been one source of Davidson's acceptance of ideal consistency conditions. In fact, Quine's translation policy itself presupposes an ideal consistency condition; in *Word and Object*, Quine often writes as if correct translation of the sentences a subject accepts must preserve ideal, rather than minimal, consistency.[2] A similar consistency assumption is implicit in Quine's holistic account of the structure of human knowledge in 'Two Dogmas of Empiricism'.[3] Davidson's transitivity condition, although generally assumed in the idealizations of decision and game theory, seems to be just false; people frequently speak and act in ways for which the best explanation is just that their preferences are inconsistent. In fact, an area of the psychology of attitudes and beliefs—consistency theory—has been devoted to study of the widespread phenomena of breakdowns in consistency.[4]

It is important to understand that inconsistencies in a belief set are not at all inexplicable. First, the logical relations among the beliefs involved in an inconsistency may be very unobvious and so

1 P. 50. See also 'Mental Events'.
2 Cambridge, 1960. For example, in ch. 2, pp. 58–61.
3 See section 6 of 'Two Dogmas of Empiricism', in *From a Logical Point of View* (Cambridge, 1961) (and also ch. 1 of *Word and Object*).
4 See W. J. McGuire's review, 'The Nature of Attitudes and Attitude Change', in G. Lindzey and E. Aronson (eds.), *Handbook of Social Psychology*, 2nd ed. (Reading, 1969).

not recognized; for example, they may be as difficult as in the Goldbach's conjecture case discussed earlier. Another important source of inconsistency, e.g., in preference transitivity, is compartmentalization—the tendency for subsets of a person's belief set to be recalled and employed in different types of situations. An inconsistency arises because the beliefs involved are unlikely to be considered contemporaneously, when the inconsistency would be more easily recognized. We shall discuss difficulty of inferences and compartmentalization of human memory later. We deal below principally with the minimal inference condition.

6. *Using minimal rationality conditions*

Having rejected the ideal inference condition, we can regard it as a provisional 'ceiling' below which minimal deductive ability must lie. The minimal inference condition remains combinatorially vague; its structure makes every intentional concept a cluster concept. The minimal inference condition by itself specifies not a 'simple defining property', but a cluster of properties for a creature's having intentional states—namely, apparently appropriate inferences from the beliefs the creature would make. With the possible exception of a 'core' of obvious inferences, any one or more of these properties may be absent, and yet the person may still qualify as having beliefs. But if all of the properties are absent, the creature does not have beliefs. As a result, the minimal conditions are generally employed probabilistically.

Dissatisfaction with the very form of the minimal rationality conditions may arise from acceptance of an oversimplified model of concepts. There is a tendency to treat all concepts as being like *bachelor* or *prime number*—that is, as defined by a single simple criterion. Also, vagueness is commonly regarded as not being present in paradigm scientific theories, such as classical mechanics or axiomatizations of normative decision theory. A simplified notion of belief, for instance that encountered in epistemic logic, where the deductive closure condition is employed, is supposed to be preferable because of its deductive manageability. In addition, one may confuse the fact that a law is not precise and quantitative (as a physical law is, supposedly) with its not having any predictive content. The indefiniteness of application of a vague term for intermediate cases restricts predictive value; however, it does not by any means eliminate it.

In addition, vagueness has advantages. Since the belief attributor, as well as his subject, is in the finitary predicament, he often cannot or ought not to obtain the evidence which would be needed to justify a perfectly precise assertion; accuracy here would be costly and unneeded. But there may be corresponding assertions employing the vague notion of minimal rationality which are justifiable by much less evidence, which it *would* be rational to obtain. Often, an assertion regarding a certain subject—e.g., the fragility of an antique chair—will only be useful if made within an interval which is too brief to permit collecting the evidence needed to justify a precise assertion; yet such an exact assertion may not be needed—e.g., as a basis for warning someone not to sit on the chair.

One may feel that, in any case, satisfaction of only the minimal inference condition would not provide as strong a basis as satisfaction of the ideal inference condition for attributing a belief-desire set to a creature. This view can be seen in Dennett's discussion of departures from ideal rationality in 'Conditions of Personhood': 'as we uncover apparent irrationality under an Intentional interpretation of an entity, our grounds for ascribing any beliefs at all wanes.'[1] Above the threshold of minimal rationality, this does not seem right; for example, failure to perform an apparently appropriate inference that is practically impossible—say, one for which a human would require more time than is available before heat-death of the universe—does not count *at all* against the person's having a belief-desire set, if he makes enough of the easier inferences from those beliefs. In effect, 'ought' seems to imply 'can' in this case, in that the person cannot be required to perform inferences which are not feasible for him. And we have a simple explanation of why the person cannot accomplish all inferences that are apparently appropriate for him: namely, that he has finite cognitive resources. Hence, a person's actions' falling short of ideal rationality need not make them in any way less intelligible to us. This leads to a more general point.

In 'Intentional Systems', Dennett may have had in mind something like minimal rationality conditions when he objected, 'If we try to fix minimum standards [of rationality] at something less than perfection, what will guide our choice?' He shortly continues, 'What rationale could we have . . . for fixing some set [of conse-

1 In A. Rorty (ed.), *The Identity of Persons* (Berkeley, 1976), p. 193; see also p. 190.

quences of a belief that are themselves believed] between the extremes and calling it *the* set for a belief (for [any given subject] S, for earthlings, or for ten-year-old girls)?' (pp. 105–6). In fact, for the minimal inference condition, this determination is based upon the cognitive psychology of the particular subject, such as theories of his deductive abilities and of his memory structure. The minimal rationality concept thus is context-sensitive. Let us briefly consider these two theories.

The content of the minimal inference condition, and in particular, the minimal consequence requirement, is considerably increased when it is employed in conjunction with a weighting of deductive tasks with respect to their feasibility for the reasoner. That is, in everyday situations the attributor of beliefs possesses an empirical theory of the difficulty of reasoning tasks for the human believer. This theory provides information about *which* inferences the believer would accomplish, namely, that the easier ones are more likely to be performed. For instance, normally, inferring $-q \rightarrow -p$ from $p \rightarrow q$ can be expected to be much easier than inferring $(x)Fx \rightarrow (x)Gx$ from $(\exists x)(y)(Fx \rightarrow Gy)$. Thus, part of the answer to our main question 'what is minimal rationality?' is provided by this *theory of feasible inferences*, which specifies more than just that some inferences must be accomplished.[1] We shall treat as an open question here whether there are particular inferences—the most 'obvious' ones, like *modus ponens*—which any creature that qualifies as having beliefs must be able to perform.

When the minimal rationality conditions are applied to human believers, a theory of human memory structure further fixes the level of rationality required for the minimal inference condition. In predicting a human being's behaviour, it is very helpful to know which beliefs will be recalled when. In particular, a useful inference is weighted in terms of whether the beliefs which are its premisses and rules are simultaneously 'activated', or being considered, at a given time by the believer. If a human is considering at one 'specious moment' his beliefs 'If I play the trumpet, the landlord will be angry', and 'I am playing the trumpet', it is at best a special aberrant case if he cannot then make the useful and very easy inference to the conclusion that the landlord will be angry. But a given inference, even one of the easiest like *modus ponens*, is evaluated as significantly more difficult if the believer has not yet

1 See my paper 'Feasible Inferences', forthcoming in *Philosophy of Science*.

'put together' the premiss-beliefs. Failure to perform the inference in the former case is worse than failure in the latter case. In effect, there are two different minimal inference conditions; the activated belief subset is subject to a more stringent inference condition than the inactive belief set. This weighting of inferences can be explained in terms of a fundamental model of *human* memory structure; it is easy to imagine believers that do not conform to this model.

The theory of feasible inferences and the theory of human memory structure are salient examples of the broad range of cognitive psychological theory in which the minimal rationality conditions are embedded. These two background theories extend further the specification of what minimal rationality is for typical human believers, and thereby help to set non-arbitrarily the 'passing grade' for minimal rationality that Dennett seemed to demand. The combinatorial vagueness of the minimal inference condition discussed earlier is correspondingly reduced. The contribution of these ancillary theories to the content of the minimal inference condition illustrates the holistic point that, by themselves, the rationality conditions can be used to make only limited predictions of actions on the basis of an attributed cognitive system.

7. *Normative conditions*

We can make one final step in identifying what minimal rationality is. We will show that the set of inferences required by the minimal inference condition is only a proper subset of the set of inferences which a believer ought to make if he is to be *pragmatically rational.* Pragmatic rationality can be understood by examining two distinct theses concerning logical compulsion, each of the form, 'In only some cases, if p implies q and a person believes p, then he must (infer and) believe q.' The *descriptive* thesis is the minimal inference condition: a believer must make some of the inferences from his beliefs which, according to his beliefs, would tend to satisfy his desires. The believer is required to make these inferences in order to be minimally rational—that is, in order to have beliefs *at all.* In a case where p implies q and a person believes p, and the inference of q from p is apparently appropriate, the descriptive thesis makes predictions that at least sometimes the person must actually perform the inference; if he generally did not, he would not be

minimally rational, and so would not qualify as believing *p* after all. The *normative* thesis is, for our purposes here: the person must make all (and only) feasible sound inferences from his beliefs which, according to his beliefs, would tend to satisfy his desires. The believer is required to make each of these inferences *if* he is to be pragmatically rational. In the case where *p* implies *q* and the person believes *p*, and the inference of *q* from *p* is feasible and apparently appropriate, the normative thesis says nothing about what the person will actually do; it says only that the person must make this inference in order to be pragmatically rational.

We can now explain the notion of pragmatic rationality in the following way. One sometimes encounters the claim 'If *p* implies *q* and a person believes *p*, then he *ought* to believe *q*'. For instance, in *Knowledge and Belief* Hintikka claims that it would be 'indefensible' or 'irrational' for someone to believe (or know) *p* and not to believe *q* here, in that he would be unreasonable and subject to criticism (pp. 29–31). However, from our earlier discussion of the finitary predicament, it is clear that it is impossible for a believer to make all of the sound inferences from the belief *p*; within limits, a believer can be rationally required only to make feasible inferences. Furthermore, only a small subset of the sound inferences which it would be practically possible for the believer to make would be positively useful for him at a given time. An inference may be sound but it may not be reasonable to make it, because it is of no foreseeable[1] value at the time and prevents the believer from doing other things which are obviously valuable at the time with his limited cognitive resources. It would be a waste of a person's time—and in some cases insane—for the person to make many of the feasible sound inferences; a person could waste his entire lifetime, probably a short one, making only such uninteresting inferences. For instance, it would not be rational for a non-suicidal creature to deduce vacuous consequences from one of his beliefs when this prevents him from making some other inference which would obviously yield information that at the time is crucial for his survival. As we saw, there could be an infinite regress of inferences involved just in deciding which inference to undertake.

1 Appropriateness here again must be evaluated relative to the agent's beliefs, not the objective facts, since it would be an unacceptably extreme idealization to assume the agent's beliefs are always correct. However, we have seen that these beliefs are subject to a consistency constraint.

Not making the vast majority of sound and feasible inferences is not irrational, it is rational.[1]

Therefore, it is true only in some cases that if p implies q and a person believes p, he ought to infer q, in that this is required for rationality. Hintikka's notion of rationality is narrow and excessively idealized, in that while a believer could be criticised for a type of epistemic inconsistency, he might nonetheless be rational when practical limitations were considered. In determining whether the person ought to make the inference of q from p in order to be pragmatically rational, we must take into account not only (i) the soundness of the inference, but also (ii) its feasibility and (iii) its apparent usefulness according to the person's beliefs and desires.

Even in those cases where the believer of p ought to infer q in order to be pragmatically rational, there is no implication that a believer of p will in fact do this. The point for pragmatic rationality is the same as for Hintikka's much stronger notion of rationality (as he applies it to knowledge): 'If [a person] knows that p and pursues the consequences of this item of knowledge far enough he will also come to know that q. Nothing is said about whether anybody will actually do so' (p. 34). What is the relation of our normative thesis to the descriptive thesis—the minimal rationality condition, which actually predicts what a believer of p will infer? The set of inferences required in a particular case for minimal rationality is only a proper subset of the set of inferences then required for pragmatic rationality. For, we do not deny that a person is rational enough to have beliefs just because he forgetfully fails sometimes to make even the most obvious, and obviously apparently useful, inferences from the beliefs. As an example, I may have established earlier $p \rightarrow q$; I may have been using it in other proofs, etc. And I may now have just proved p, and have been using it subsequently, etc. And it may be that I must see that q is true before some other desired proof can be completed, but I may not have recognized this yet. Nonetheless, I can still qualify as believing p.

Thus, it is a fact of our actual belief-attributing practice that minimal rationality is weaker than even pragmatic rationality.

1 Simon (op. cit.) is well known in decision theory for having made a similar point, that an agent ought only to attempt to 'satisfice', rather than 'maximize'. (To reply that the principle of maximization is also to be applied to choices about how to expend cognititive resources will simply again introduce infinite regress.)

Furthermore, one can argue that a satisfactory descriptive cognitive theory should employ a rationality condition that requires less than 'perfect' pragmatic rationality, as we argued that such a minimal rationality condition should be weaker than ideal rationality: Humans and other intelligent creatures are generally at least moderately inefficient, forgetful, and careless; using the above normative condition as a descriptive condition would again exclude, although not so extremely as the ideal conditions, these creatures from having a cognitive system. This would be undesirable because it would prevent the observer from taking advantage of most opportunities for predicting behaviour on the basis of an intentional theory, which is typically the only feasible means of doing so. In this way, the actual is not, and ought not to be, the ideal; minimal rationality should not be perfect pragmatic rationality. Not surprisingly, ideal rationality conditions, and hence the impossibility of a predictive intentional theory, gain plausibility when descriptive minimal rationality conditions are not distinguished from normative rationality conditions; this is particularly noticeable in Dennett's and Davidson's accounts.

It is the concept of perfect pragmatic rationality specified by the normative condition above which is needed for a 'naturalized epistemology' that takes account of the psychology—for example, the limitations, current beliefs, and goals—of the knower. This is especially clear in a recent paper by Alvin Goldman, in which Goldman proposes 'a reorientation of epistemology' in the form of an enterprise of *epistemics*.[1] Epistemics 'would seek to regulate or guide our intellectual activities', and would recognize that such 'advice in matters intellectual . . . should take account of the agent's capacities' (pp. 509–510). Such a programme rejects the almost exclusive concern of traditional epistemology with ideal agents of virtually unlimited cognitive resources.

After examining a number of rationality conditions, we have a first approximation of a minimal condition on the deductive ability required of a belief system. Figure 1 shows the overall scheme for the various rationality concepts. We have been principally concerned with the 'ceiling' on required rationality rather than the

1 'Epistemics: The Regulative Theory of Cognition', *Journal of Philosophy* (1978) (Goldman (p. 510 and p. 514) and I have independently arrived at some similar observations). Two important earlier proposals were Quine's 'Epistemology Naturalized', in *Ontological Relativity and Other Essays* (New York, 1969), and Donald T. Campbell's 'Evolutionary Epistemology', in P. A. Schilpp (ed.), *The Philosophy of Karl Popper*, vol. I (LaSalle, 1974).

'floor', because the ceiling is generally ignored. The ongoing project remains of further characterizing minimal rationality—for instance, by investigating ancillary theories employed in attributing beliefs. The sketch of a conception of minimal rationality we now have is by itself a step toward explaining our considerable success

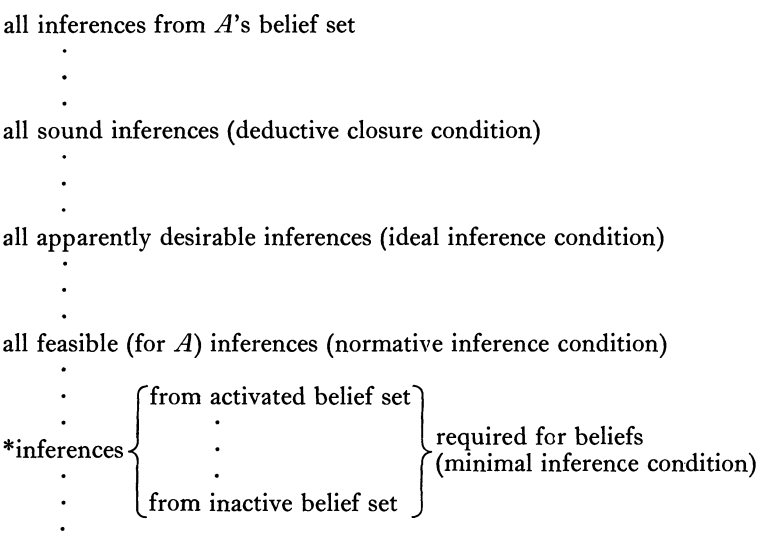

Figure 1. Partial specification of minimal deductive ability.

as predictors of behaviour, e.g., in the Holmes story with which we began; we can see in this respect how everyday psychological practice can be so sophisticated and robust. Furthermore, what we have found for common-sense psychology, however 'primitive', should apply to a manageable and predictive 'scientific' psychology: It seems that any theory of belief which is to satisfy the fundamental constraints of having significant empirical content, applying to finite creatures more than 'in principle', and being applicable by finite creatures must include the basic principle that a believer has some, but not ideal, logical ability. While the minimal rationality conditions on belief are not usefully regarded as 'definitional', they must also be distinguished from mere empirical generalizations about human psychology, such as a claim concerning our short-

term memory capacity; we have seen that the minimal rationality conditions have a centrality in a theory of belief such that they could not be rejected on the basis of just some putative counter-examples. The important point concerning the possibility and nature of a cognitive psychology is that the minimal rationality conditions seem to be indispensable in this way for any satisfactory cognitive theory.

TUFTS UNIVERSITY