FEASIBLE INFERENCES*

CHRISTOPHER CHERNIAK†

Department of Philosophy
Tufts University

A philosophically important but largely overlooked cognitive theory is examined, one that provides information on which inferences an agent will make from his beliefs. Such a theory of feasible inferences is indispensable in a complete cognitive psychology, in particular, for predicting the agent's actions on the basis of rationality conditions and attributed beliefs and desires. However, very little of the feasibility theory which applies to a typical human being can be shown a priori to apply to all agents. The logical competence required of a rational agent seems to have a cluster structure: it cannot be the case that an agent is able to make no inferences, but an agent can be unable to make any particular one.

1. Role of a Theory of Feasibility. One can argue that the actions of an agent can be predicted on the basis of an attribution to him of a system of beliefs and desires. A major element of any such predictive account of psychological states seems to be that the belief-desire set is subject to minimal, as opposed to ideal, rationality conditions.1 Furthermore, such minimal rationality conditions seem to be an element of our concept of a person; satisfying them is a necessary condition for qualifying as a person. Indeed, this type of a concept of a "minimal person" seems in turn to be required for an adequate functionalist account of psychological states, as well as, more generally, for a predictive cognitive theory. If one restricts consideration only to ideal rationality conditions (for instance, deductive closure of the belief set), as Hintikka (1962), Dennett (1971), Davidson (1976), and many others do, then, since nothing remotely satisfies such a theory, there cannot be an interestingly applicable and predictive cognitive theory.

A sound motivation for idealizations is that they yield simplified and manageable theories. However, there can be different degrees

*Received November 1979; revised July 1980.
†I wish to thank Charles Chihara, William Craig, Daniel Dennett, and Barry Stroud for their generous help. Drafts of this paper were read at the Logic and Methodology Colloquium of the University of California, Berkeley, 1976, and at the 1978 meeting of the Society for Philosophy and Psychology in Boston. Much of the material in this paper was presented, in somewhat different form, in Cherniak (1977).
1I discuss the nature of, and need for, minimal rationality conditions in a predictive cognitive theory in Cherniak (1981).

Copyright © 1981 by the Philosophy of Science Association.
of idealization of a theory; in this way, one can exchange applicability for manageability. The theory discussed here can be regarded as an attempt to obtain some significant empirical applicability in exchange for a more complex theory. An important respect in which the discussion will remain idealized is that I shall deal primarily with believers as "sentence processors." That is, until the last sections of this paper, I shall be representing an agent's beliefs as a set of sentences, and an inference from those beliefs as the addition of a sentence to that set\(^2\) (the process need not be conscious or deliberate).

What logical abilities must a predictive cognitive theory attribute to an agent? The assumption that an agent can make very complex inferences from his beliefs is crucial in our attributions of psychological states in everyday situations. The presupposition is also involved in the motions of legal responsibility and moral culpability. It has a similarly central, if often tacit, role in historical explanation and economic theory. The assumption also is required by philosophical accounts of meaning in terms of speakers' and hearers' intentions. The rationality condition with which I shall be principally dealing is a deductive inference condition:

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

An inference is apparently appropriate if, according to the agent's beliefs, it would tend to satisfy his desires.

As I explain in Cherniak (1981), the argument for such a minimal rationality condition on logical ability proceeds by exhaustion of a trichotomy. Briefly:

1. In order for a cognitive theory to have any empirical content (i.e., if one is to be able to predict an agent's behavior on the basis of the theory and attributed beliefs and desires), it cannot be the case that according to that theory the agent can perform none of the inferences that are apparently useful for him.

2. But, in order for a belief-desire attribution actually to be applicable to finite creatures, the agent cannot be required to perform all such inferences.

3. Hence, the believer must perform some of those inferences.

Such minimal rationality conditions are embedded in a broad range of ancillary cognitive psychological theory. By themselves, these

\(^2\)This simplified model of inference therefore can be contrasted with, for instance, the account in Harman (1973).
rationality conditions can be used to make only very limited predictions of actions on the basis of an attributed belief-desire set. For instance, the above inference condition does not specify how long it would take an agent to perform a useful inference. If the attributor makes the farfetched assumption that inferences may take the agent up to some colossal amount of time, then the rationality conditions will not have content as a practical matter — e.g., as a basis for expectations that the agent will decide on appropriate actions. The need for such ancillary theory is not peculiar to psychology; physical theories similarly require a broad range of background presuppositions (e.g., classical mechanics depends in this way on measurement theory). I shall not attempt to enumerate exhaustively the entire range of background theories typically taken for granted in using rationality conditions. I shall deal with only one of the philosophically more interesting ancillary theories.

In predicting an agent's actions on the basis of his current beliefs and desires and the above inference condition (among other rationality conditions), the question must be answered, which deductive inferences from the beliefs are most likely to occur? This question has two parts: first, what inferences are most likely to be undertaken (not necessarily consciously); and second, which of these are most likely to be accomplished? I am concerned here with the latter question, the feasibility of inferences once they have been undertaken, as opposed to the prior heuristic task of determining that particular deductive tasks will be appropriate or valuable. Without an extensive theory of the difficulty of different inferences, which provides us with information on which ones will be accomplished (under given conditions), the predictive value of any cognitive theory of beliefs and desires would be severely limited. I shall refer to this theory of the difficulty of inferences for an agent as a theory of feasible inferences. I shall be concerned only with inferences of classical first-order logic; feasibility theories for second-order and inductive logic are also important and interesting elements of a complete cognitive theory. An easy inference is taken here to be one which would be reliably accomplished, given that it had been undertaken.

Traditionally, "self-evident" or "clear and distinct" truths of reason, which are known directly and apparently reliably "by intuition," have been distinguished from truths that are not intuitively obvious. This distinction can be generalized in terms of the notion of an ordering of inferences with respect to relative difficulty. A theory of feasibility will include such an ordering. Thus, other things being equal, inferring \( p \rightarrow q \) from \( -q \rightarrow -p \) typically is easier (and hence more likely) than inferring \( (x)Fx \rightarrow (x)Gx \) from \( (\exists x)(y)(Fx \rightarrow Gy) \), and the latter
task is much easier than one as difficult as determining that the axiom of choice is independent of the other axioms of set theory. An even more difficult inference would be one which required more space and time than is available in the galaxy before heat-death. Of course, inferences involving sentences of the same logical form can differ in difficulty (e.g., two instances of *modus ponens* can so differ when the consequents of the two conditionals differ in complexity). The feasibility ordering can, as a useful simplification, be roughly described as a well-ordering of equivalence classes of inferences. In particular, some inferences are (*ceteris paribus*) the least difficult, such as, typically, inferring \( p \) from \( p \land q \); however, no inference is more difficult than all others.

2. Universality. Let us consider the feasibility theory which in fact applies to some normal human being at a particular moment. The main question of this paper concerns the universality of this theory. One version of this question is: what logical abilities described by this theory must be possessed by any creature that is an agent, and in particular, satisfies a minimal rationality condition on deductive abilities? As mentioned above, minimal rationality conditions seem required for *any* cognitive theory that (a) has predictive content and (b) is significantly applicable to creatures of limited resources. This centrality of the rationality conditions must be distinguished from a true but clearly contingent empirical psychological generalization, such as the finding that a normal human can retain no more than about six unrelated "meaningful units" (e.g., random digits) in short term memory at a time. (It should similarly be distinguished from more widely applicable assertions regarding practical learnability and usability of particular types of language or logic.)

Our question here is, what features of a theory of feasibility have a status like that of the rationality conditions—i.e., a similar universality—and which features have only the status of a generalization about human psychology, or an even more limited applicability? The difficulty of an inference is dependent not only on the inference itself, but

---

3I employ lower-case italics 'p', 'q', and 'r' as variables ranging over sentences. Otherwise, logical notation here is that of language in Mates (1972). Single quotation marks will be used to name the expression occurring within the quotation marks; double quotation marks will be used as "scare quotes" and for quotation of an author's work. For our purposes, two kinds of belief attribution can be briefly distinguished: On the one hand, the assertion 'John believes \( p \lor \neg p \)' says that, for a sentence \( p \), John believes the disjunction \( p \lor \neg p \); for example, John might believe the English sentence 'It's raining or it isn't raining'. On the other hand, the assertion 'John believes \( 'p \lor \neg p' \)' says that John believes (or accepts) the logical law of the excluded middle, and not just instances of the law.
also on the conditions under which it is performed. Feasibility orderings differ considerably among humans (e.g., depending on training in formal logic, problem-solving "set", etc.); indeed they can fluctuate for a single individual from one minute to the next, for instance, as he learns a new deductive strategy. It is false that the difficulty of any given inference is always the same for all reasoners, or even all normal humans. Consequently, the difficulty of an inference cannot correspond to any syntactical feature which is intrinsic to the inference, in isolation from the reasoning psychology of the deducer.4

There are several important empirical questions in the field of psychology of reasoning regarding the feasibility orderings of human beings. One question is whether certain logical operations or inferences are easier for all "naive" humans with no training in formal logic; and if so, which ones are natural or "intuitive" in this way? Authors of elementary logic texts often claim that they have chosen the inference rules and axioms of their deductive systems to be 'convenient', 'simple', etc. A question for empirical research is, which inference rules are in fact most easily learned and efficient for the purposes of formal deductions? One important impediment in empirical psychology of logic seems to be that there is uncertainty regarding the status of such questions—that is, regarding which parts of a theory of feasibility even need to be determined by empirical research, as opposed to some a priori transcendental argument.

The principal claim of this paper is that there is greater latitude for such empirical investigation than has generally been granted. As we will see, there seems to have been a tendency in philosophy to treat contingent facts of common-sense psychology of reasoning as a priori universally applicable truths. In fact, it seems that little of a typical human feasibility ordering can be shown a priori to apply to all agents; hence, we must determine most features of the feasibility orderings that apply to human beings by empirical research.

3. Alternative Feasibility Orderings. One source of uncertainty regarding whether assertions about the logical abilities of, say, a rational

4Hintikka seems to suppose, to the contrary, that there is some objective measure of the difficulty of an inference which must apply to all reasoners. Hintikka says, "if this consequence-relation [of p to what I know] is a distant one, I may fail to know, in a perfectly good sense, that p is the case, for I may fail to see that p follows from what I know" (1962, p. 30; also p. 35). The suggestion seems to be that the "remoteness" of the logical relation between two statements is independent of, for instance, the deductive system used by the reasoner. See also Hintikka (1970), especially p. 147.
FEASIBLE INFERENCES

253

creature are a part of empirical psychology is the fact that features of a feasibility theory vary in their status. In particular, two alternative theories of feasibility are ruled out as more than just a matter of an empirical generalization. One of these alternatives is a universal feasibility theory, according to which every inference is maximally feasible, that is, always accomplished. With such an extremely idealized account of the logical abilities of an agent, no creature with fixed resources could qualify as having beliefs. The other extreme alternative is a null feasibility theory, which consists of the assertion that every inference is maximally unfeasible, that is, never accomplished. The null theory just says that all creatures that conform to it are entirely logically incompetent, and so that no such creature could have a cognitive system.

Thus, the denials of the universal and null feasibility theories, and of approximations to them, are as universally applicable as the minimal rationality conditions on beliefs themselves. No empirical investigation is needed to justify rejecting the universal and null theories. The assertion of the universal theory that agents are ideal deducers generates so many incorrect predictions for any belief attribution, that it thereby effectively excludes human beings (and virtually any finite creature) from having a cognitive system; the null theory directly excludes any creature from ever having a cognitive system. We mentioned earlier that feasibility orderings can and do often differ in “minor details”, to some degree, from one human to another. But we must now consider the possibility of certain alternative feasibility orderings which differ from, for example, some normal logically naive human’s more than these common moderate variations, but less than the radically different universal and null orderings. Could there be an agent who conformed to one of these intermediate alternative feasibility orderings? These are the most philosophically important cases. According to these alternative theories, the inferences which are most easy for us, such as modus ponens, are evaluated as being of greater levels of difficulty, i.e., as not being easy.

Let us consider the set consisting of all inferences from the least difficult for this particular normal human being through the least difficult ones which are practically unfeasible. One intermediate alternative to this person’s feasibility ordering would be one in which this initial segment of the actual ordering was inverted: the easiest inferences became the most difficult within this initial segment, the most difficult inferences became the easiest in the new ordering, and so forth. According to the feasibility theory describing this alternative ordering, inferring \( q \) from \( p \rightarrow q \) and \( p \) would not be possible, while performing an inference of the level of difficulty (for
the original human agent) of determining the independence of the axiom of choice would be an easy task, performed reliably and without prolonged investigation. Supposedly, failure at the former task would be common, while failure at the latter task would be rare.

Thus, at the very least, typical humans would be doing remarkably and inexplicably well at the more "difficult" inferences, such as inferring \( q \) from \( p \rightarrow q \) and \( p \), but they would always be failing at the "easier" inferences, such as the axiom of choice example. In this way, the alternative feasibility theory clearly would generate many incorrect predictions regarding the reasoning behavior of these people; anomalous failures and successes would be the rule. And so, in conjunction with the inference condition, there would be a strong empirical disconfirmation of any correct attribution of an explicit belief set to one of these people.\(^5\) If a typical human's beliefs included 'If it rains, the dam will break' and 'It's raining', the alternative feasibility theory would imply that it was unlikely that the person would infer that the dam will break and act on this new belief (e.g., by seeking high ground); but in fact this inference and the corresponding action would be very likely. And according to the alternative feasibility theory, we would expect particular actions that would be based on inferences that in fact would be very unlikely to occur; the actions would not occur because the required inferences from the relevant beliefs would not occur, and so again the correct belief-attribution would be disconfirmed. Therefore, as long as, say, a Martian belief-attributor accepted an alternative feasibility theory of this kind, where a large initial segment of the feasibility ordering was inverted, a "normal" human would not, according to the Martian's theory, qualify as having a cognitive system. This is not surprising, given the important role of a theory of feasibility in a complete cognitive theory, since by hypothesis this alternative theory was false—it didn't apply to the subjects considered here.

The next question concerns the possibility of an agent whose logical abilities conform to this alternative feasibility ordering; in particular, could there be a creature which satisfied the rationality conditions but conformed to the alternative ordering? This would be a creature which, in relation to the feasibility theory for a typical human, would in general not be able to perform any very easy inferences (for example, inferring \( q \) from \( p \rightarrow q \) and \( p \)), but would be able to perform many very difficult inferences like the axiom of choice example quickly and reliably, with a similar inversion of the intermediate cases.

\(^5\)See section 7 for a qualification regarding nonverbal beliefs.
4. "Unnatural" Deductive Systems. Certainly it is a fact of human psychology at least as basic as our having a short term memory capacity of about six "chunks", or even a distinct short term memory at all, that when we are naive about formal logic our deductive abilities do not conform to this inverted feasibility ordering. However, we must now consider some arguments for the stronger conclusion that a creature with anything like an inverted feasibility ordering is ruled out as more than just an empirical matter of human psychology. The first argument is that it would be much more difficult to explain how a creature could succeed in performing typically complex inferences (like inferring \((x)Fx \rightarrow (x)Gx\) from \((\exists x)(y)(Fx \rightarrow Gy)\)) but consistently fail at simple ones like modus ponens, than to explain how a creature could have a larger short term memory than we do.

The mystery for the creature with the inverted feasibility ordering is, if this creature cannot perform any simple inference like modus ponens, then how is he able to perform much more complicated inferences? This argument may be assuming that, in explaining the creature's logical ability, we must employ a "principle of axiomatic method":

A deducer—or, a deducer that is a rational agent—must generally perform complex deductive tasks by performing a sequence of simpler deductive tasks.

The more complex tasks would be taken here to be the more difficult inferences of a typical human being's feasibility ordering (like the above quantification theory example), the simpler tasks to be the easier ones in the ordering (like modus ponens). So, this principle requires that an agent's problem-solving process for deductive tasks be somewhat like the way in which a formal deductive system is used: simpler theorems are proven by employing elementary rules, and sometimes axioms, which are accepted as sound; other theorems are in turn proven by using these previously proven theorems, or by proving appropriate additional lemmas.

A creature with an inverted feasibility ordering seems to violate this requirement, since it is least likely to infer the "axioms" and simpler "theorems", and so will not use them in inferring the more complex "theorems", yet it can still infer these complex "theorems"—and infer them much more easily. The "principle of axiomatic method" does seem to be part of our common-sense psychology of problem-solving in general, and of human deductive reasoning in particular—that is, the sort of deductive reasoning that even humans who have never encountered a formal system perform.
We do explain normal human performance of complex inferences as proceeding by means of performance of simpler ones.

An adequate reply to this kind of argument is that it is easy to imagine a creature with psychological processes which would result in its having at least a very scrambled feasibility ordering with respect to the "normal" one, but which would still conform to this axiomatic methodology. Let us consider a hypothetical creature that exclusively uses a deductive system of axioms and inference rules in making inferences from its beliefs when it decides on actions. This system would consist of inference rules and axioms which would be unintuitive and complex for us; for example, an instance of one of the axioms might be, 

\[(\exists x)Fx \rightarrow p \leftrightarrow (x)(Fx \rightarrow p).\]

It is clear that the inferences corresponding to each of these rules and axioms, which would be difficult for us, would be trivial for this creature. And if this deductive system was suitably designed, inferences which were easy for us would be possible but difficult using this system—involve many steps, special strategies, etc. Indeed, this latter feature is very common in the formal deductive systems presented in elementary logic texts, despite efforts to the contrary: some of the inferences which are in fact the most trivial for humans without training in formal logic are much more difficult than some inferences which are normally much less intuitive. (For example, inferring \(q \land p\) from \(p \land q\) is much harder than inferring \(q \rightarrow (p \rightarrow r)\) from \(p \rightarrow (q \rightarrow r)\) in the system in Mates (1972).) In this way, the logical ability of a creature that used this deductive system would approach an inverted feasibility ordering, and yet the creature would perform "easy" (for us) tasks by means of "difficult" (for us) tasks. The argument against the possibility of a creature with an inverted feasibility ordering thus does not seem promising.

For a second, even stronger reply to the above argument, we must first consider another creature that performs all of the inferences involved in selecting its desired actions by means of (a limit case of) a deductive system. This creature would employ a list of theorems of logic (or, substitution instances of such theorems), containing a finite number of items, but including all theorems up to some colossal level of logical complexity. The items might be lexicographically ordered so that they could be searched efficiently and rapidly. With this list, the creature could perform any inference that an intelligent human agent could; this ability would account for all relevant actual or possible deductive tasks. Given any question of whether \(p\) is a

---

6This axiom corresponds to one of the basic inference rules of, among others, the deductive system of Kleene (1967, p. 107).
logical consequence of a set (possibly empty) of premises, the creature would form the associated conditional and search the theorem list. If the conditional was on the list, the creature would infer \( p \) from the premises; if it was not on the list (at the appropriate location in the ordering), the creature would not make the inference. Thus, this creature employs an "unnatural" deductive system consisting of an enormous axiom set and a single ("theorem") inference rule.

This system cannot be complete. But it would be practically adequate, in that the creature using it could have the logical ability needed to satisfy the inference condition for having beliefs. Some creature of this kind could perform any deduction which a normal intelligent human could; in fact, for any given creature with fixed finite resources, there would be some creature of this "unnatural" type with the same deductive abilities. However, it is clear that this creature's "all or nothing" feasibility ordering is not anything like the normal feasibility ordering for humans. The creature also does not conform to the above principle of axiomatic method. The creature would never follow the procedure of proving complex theorems by first proving simpler ones; indeed, the creature would never prove any theorems by first proving other theorems. (Note that assuming some inferences must be made even in using the theorem list will also result in a type of regress similar to the Tortoise and Achilles [see Stroud 1979] for our own ordinary reasoning.)

Now, if all theorems of less than a given level of logical complexity (e.g., of less than a particular large number of logical words in the corresponding schemata) are deleted from this creature's theorem dictionary, its logical ability will approximate (with allowance for its "all or nothing" character) the inversion of a normal human feasibility ordering. Thus, a creature that violated the principle of axiomatic method and conformed to a very different feasibility ordering, such as the original creature above or this modification of it, is in fact possible.

The next question is whether such a creature could qualify as an agent—specifically, satisfy the rationality conditions. These creatures' procedure in performing a given deductive task might seem inefficient and heuristically obtuse, in that they could never use in accomplishing the given task any results that they had previously obtained. (Perhaps these creatures could only survive in environments rather different from our typical terrestrial one.) However, it should be noted first that the behavior of these creatures is not entirely bizarre; it differs only in degree from that of normal human beings. Humans sometimes, although not always, similarly use lists of theorems established by themselves or others, and use proof procedures. And humans often
just forget, in performing an inference, to use results they have established earlier, even when the earlier results are obviously useful; consequently, there is an inefficient repetition of effort.

It seems that a creature of the above kind could accomplish enough of the inferences it undertook to qualify as having beliefs. For, such a creature could have sufficient cognitive abilities—in particular, be so fast in searching its theorem dictionary—that the inherent inefficiency of its deductive procedures would not, by itself, rule out its having a cognitive system. It also seems that a creature that searched its theorems effectively enough would satisfy any requirement on heuristic ability; it would undertake enough of the inferences which were apparently useful. Narrowly considered, a creature of this type would seem in one area to be heuristically imbecilic, in that the most obviously useful inferences (by any standard) would not be undertaken when their apparent usefulness was in accomplishing another deductive task. But if this creature's peculiar deductive process was rapid enough, there might in fact be no gain in efficiency to be had from its instead deducing in the normal way that we do.

The inference condition is only a necessary condition, on logical abilities, for having a cognitive system. To the extent that we regard satisfaction of the inference condition as possession of all of the logical ability required for having beliefs, i.e., as a sufficient condition for being logically competent to have beliefs, satisfaction of the inference condition should permit the creature's satisfaction of other rationality conditions on having beliefs (including a condition requiring maintenance of belief-set consistency). The above discussion then shows at least that a creature that is logically competent to have a cognitive system does not have to conform to the axiomatic method.

One response to finding that the above creature with its theorem dictionary could satisfy the rationality conditions may well be that this suggests the rationality conditions are too weak; a complete or adequately rich cognitive theory would exclude such a creature. Our strategy in dealing with this issue of completeness will be to consider below several questions which arise concerning a creature with a radically different feasibility ordering. This procedure cannot decisively establish that such a creature could have a cognitive system; only a complete cognitive theory could do that. But this strategy will at least show that certain philosophically important theses underlying objections against the possibility of such an agent are mistaken. A question for further research will remain.

Let us first provisionally draw the philosophical moral of the above discussion of creatures with inverted feasibility orderings. One point is that we must distinguish a psychology that seems radically different
FEASIBLE INFERENCES

from ours, namely, that of such a creature, from a logically impossible or inexplicable psychology. A creature whose deductive processes involved the use of a suitable theorem dictionary would, without any unaccountable mystery, manifest overt abilities conforming to the inverted feasibility ordering. It seems at this point that, so far as rationality is concerned, an agent must only be an adequate logician, not "the right kind" of logician; beyond empirical generalizations about human psychology, there is no particular right kind of logician. I referred at the beginning of the paper to the cluster structure of the concept of belief; it is apparent here for the logical abilities required by the minimal rationality conditions. It cannot be the case that an agent can make none of the inferences from his beliefs, but the agent can be incapable of making any particular inference.

5. Constitutive Inferences. The next type of question concerning agents with deductive abilities radically differing from our own can be introduced by the following argument: If a creature denies or suspends belief in an obvious logical law like \( -(p \& -p) \), then he does not understand the logical constants occurring in the law. One question this argument raises is, what accounts for the alleged special status of obvious laws like \( -(p \& -p) \)? Normal human beings, of course, often accept some obvious laws and deny or suspend judgement on many complicated ones, such as \( (\exists y)(x)(Fy \lor (Fx \rightarrow p)) \). We also have seen that there apparently can be deducers whose logical abilities are the mirror image of ours, who accept many of the theorems we do not, and deny or suspend judgement on many of the theorems we accept. Why would not their acceptance of those theorems constitute as good an understanding of the logical words involved as our acceptance of other theorems? However, the argument here can be reinforced by citing the difficulties in translating and identifying the logical constants for deducers with supposedly radically different logical abilities. The conclusion of the argument would still be that being able to perform certain "easy" inferences is constitutive of understanding a given logical constant, and so is necessary for believing any sentence containing that constant.

Let us consider an instance of this argument, the claim that for the connective 'or,' being able to infer \( p \lor q \) from \( p \) ("'or'-introduction") and to infer \( q \) from \( p \lor q \) and \( -p \) ("disjunctive syllogism") are required in this way: in the simplest situation, if the deducer would assent to the premises but not generally to the conclusions of these inferences, there could be no basis for determining that the deducer meant by 'v' or 'or' what we mean by 'or', as opposed to, e.g., what we mean by 'and'. Whatever inference abilities remained could not be
sufficient to distinguish ‘or’ from ‘and’ for this creature. For example, suppose (I) the hypothesis for some creature’s inference behavior was that ‘*’ was to be translated as ‘or’, although the creature did not infer \( p * q \) from \( p \) or infer \( q \) from \( p * q \) and \( -p \) as above for ‘v’ (see Fig. 1). And suppose (II) the evidence was that whenever the creature assented to \( p * q \) it would assent to \( q * p \) but not necessarily to \( r \)—i.e., according to the hypothesis, the creature still had the logical ability to infer \( q v p \) soundly from \( p v q \) and not to infer \( r \) unsoundly from \( p v q \). This inference behavior would not provide any evidence that ‘*’ was to be translated as ‘or’ rather than ‘and’, because the inference of \( q \& p \) from \( p \& q \) is also sound, and the inference of \( r \) from \( p \& q \) unsound. The creature’s inference behavior equally supports the hypothesis that ‘*’ should be translated as ‘and’.

The reply to this argument is that we must recognize that the inability to make “easy” sound inferences like inferring \( p v q \) from \( p \) does not imply an inability to make other “harder” inferences; then it is clear that the argument does not show that for a deducer who does not make the two allegedly “constitutive” inferences for ‘or’, there can be no basis for determining that the deducer means by ‘or’ what we mean by ‘or’, as opposed to what we mean by ‘and’. For, a creature that cannot make these two inferences may have enough other inference ability so that we can distinguish ‘or’ from ‘and’ for

<table>
<thead>
<tr>
<th>Inferences involving ‘*’</th>
<th>the creature makes</th>
<th>the creature does not make</th>
</tr>
</thead>
<tbody>
<tr>
<td>I) ( p )</td>
<td></td>
<td>( p * q )</td>
</tr>
<tr>
<td>( -p )</td>
<td>( p * q )</td>
<td></td>
</tr>
<tr>
<td>( q )</td>
<td>( p )</td>
<td></td>
</tr>
<tr>
<td>II) ( p * q )</td>
<td>( p * q )</td>
<td>( r )</td>
</tr>
<tr>
<td>( q * p )</td>
<td>( p )</td>
<td></td>
</tr>
<tr>
<td>III) ( -(-p * q) )</td>
<td>( p * q )</td>
<td>( p )</td>
</tr>
</tbody>
</table>

Fig. 1.
FEASIBLE INFERENCES

261

this creature. No question-begging is involved here; just given, for example, an object that behaved in the way we have described earlier as "having an inverted feasibility ordering", we could go on to establish the translation of logical constants for it.

Let us assume that we have established the translation of negation for this creature that seems to use "*" as a connective, say, that the creature means by "-'" what we mean by it. Now, suppose (III) that whenever the creature assents to \(-(-p \land q)\) he would assent to \(p \land q\) (and whenever he assents to \(-p \land q\) he would assent to \((-p \land s) \land (-r \land q)\), etc.). So long as we are assuming the creature has logical ability, this is evidence that "*" is to be translated as 'v' rather than '&', because the inference of \(p \lor q\) from \(-(-p \lor q)\) is sound, while the inference of \(p \land q\) from \(-(-p \land q)\) is not (similarly for the other inference). Conversely, suppose that it is not true that whenever the creature assents to \(p \land q\) he would assent to \(p\) (and that it is not true that whenever he assents to \(p \land q\) he would assent to \(-(-p \land q)\), etc.). This is further evidence that "*" is not to be translated as '&' and should be translated as 'v', because the inference of \(p\) from \(p \lor q\) is not sound, while the inference of \(p\) from \(p \land q\) is (similarly for the other inference).

In the simplified translation situation we are considering, two types of inference ability provide a basis for determining that a deducer understands and uses a given logical constant: First, there are the sound inferences involving that constant which he will make, but which are not sound (and which he would not generally make) when another constant is substituted. Second, there are the unsound inferences involving that constant which he will not make, but which are sound (and which he would still make) when another constant is substituted. On the basis of these two types of evidence, we could similarly establish that the creature was using negation; and in this way we could also determine that the creature was using the universal quantifier. Even if the creature used no other logical constants, he would then be able to express any sentence in which other constants occurred.

The failure of this argument against the possibility of a deducer that did not make the so-called "constitutive" inferences suggests that a cluster structure is again involved—here, in the identity of the logical constants. The meaning of the constants is, at least partly, determined holistically by the entire range of accepted laws and inferences in which they occur; this claim is just a special case of a holistic account of meaning.\(^7\) Quine (1970) says that the identity of a logical constant is determined by the accepted laws (more generally,

\(^7\) Such as was proposed in the last section of Quine (1963).
let us say inferences) in which it occurs. And Quine correctly points out that if for someone "all the laws which have up to now been taken to govern alternation were made to govern conjunction instead, and vice versa," the person's 'or' would merely become our conjunction, and vice versa (Quine 1970, p. 81). But Quine does not distinguish between rejecting all of the laws in which a given logical constant occurs and rejecting some of them.

We have found that, while a deducer cannot be incapable of making all of the inferences governing a given constant, he can be incapable of making any particular inference involving the constant (if he can make enough others) without changing the identity, and our translation, of the constant. However farfetched the above creatures with deviant feasibility orderings may seem, and however unlikely it may be that these cases might arise in practical situations, they are important in showing what is necessary for the identity of logical constants. If we do not adopt the "cluster structure" view, then we are faced with the problem of finding a satisfactory objective basis for the special status of "obvious" sound inferences or logical truths, as opposed to any other logical truths.

6. 'Believing That ______'. One may feel that the account thus far of agents with radically different logical abilities from ours has been too syntactical or formalistic. We have been principally concerned with the expression 'A believes '______''; that is, our account of the theory of feasibility has dealt with beliefs in sentences rather than, say, propositions. The criticism that the account here is excessively verbalistic suggests that we should now also examine the expression 'A believes that ______'; for, one might argue that creatures with radically different logical abilities from ours only seem possible so long as we deal solely with a creature manipulating sentences, and so do not consider the latter expression.

This latter construction is of course referentially opaque—that is, intersubstitution of coreferring expressions after 'that' does not always preserve truth of the entire belief-sentence. But is it "intensionally opaque"? Intersubstitution of synonymous expressions is sometimes alleged to preserve truth in this case; if a person believes that John is a bachelor, then the person believes that John is an unmarried male. Various "closely" logically equivalent sentences are commonly regarded as synonymous (or as paraphrases, or as expressing the

---

8 See Hilary Putnam's important qualification of this assertion in section 3 of Putnam (1975/6). The point remains that accepting certain logical laws is necessary, even if not sufficient, for the identity of a constant.
same proposition), such as $p \land q$ and $q \land p$, or $p$ and $\neg p$. So, by the intensional transparency claim of intersubstitutability of synonyms in belief contexts, for instance, if a person believes that $a = b$ and $c = d$, then he believes that $c = d$ and $a = b$.

Now, the objection against the possibility of a creature with, for example, an inverted feasibility ordering, is: if it is true that such a creature believes that $a = b$ and $c = d$, then by the intensional transparency thesis the creature must also believe that $c = d$ and $a = b$. It turns out, then, that in terms of the 'believes that _____' construction, the creature cannot have a deviant feasibility ordering, where this inference, which is obvious for us, is difficult (i.e., not always performed). But according to the above account of the theory of feasibility, the creature (if he was a suitable English speaker) could believe the sentence 'a = b and c = d' while not (inferring and) believing 'c = d and a = b.' The objection here suggests that this disparity between the creature believing that $c = d$ and $a = b$, but not believing the sentence 'c = d and a = b' must be resolved by abandoning the normal homophonic translation of the sentence. If a creature allegedly has very different logical abilities from ours, so that equivalences which are close for us are remote for him, then this requires us to alter our translations of his sentences so that his abilities conform to ours; otherwise the above disparity arises. In this situation, the 'believes '_____'' construction must be used with caution.

The problem for the intensional transparency thesis which is relevant here arises when we ask: for whom must the intersubstitutable expressions supposedly be synonymous? Not for the agent; in the above 'bachelor' example the agent might not speak English, and so not know that 'bachelor' and 'unmarried male' are synonymous. Therefore, the intersubstitutable expressions must be synonymous for the attributor of the beliefs. If we do not consider a creature's psychology, all logical equivalences have the same status—those that are close for us, and those that are remote for us. The following counterexample to the intensional transparency thesis is based on the fact that: that thesis permits intersubstitution of expressions which are synonyms for the attributor; but what are synonyms—and, in particular, closely equivalent expressions—for the attributor may not be synonyms, even in translation, for the agent.

We discussed earlier a case where a creature performed all of his inferences by means of a deductive system, and where this system, although sound and complete, contained only axioms which we (as normal humans without training in logic) would find very counterintuitive (such as $((x)Fx \rightarrow (x)Gx) \leftrightarrow (\exists x)(y)(Fx \rightarrow Gy)'$). There are
deductive systems of this kind in which it would be as difficult for an agent to determine the equivalence of \( p \& q \) and \( q \& p \) as it is for normal human beings to establish the equivalence of \((x)Fx \rightarrow (x)Gx\) and \((\exists x)(y)(Fx \rightarrow Gy)\) (and vice versa); indeed, this holds to a significant degree for deductive systems in logic textbooks that are intended to be "intuitive" for normal humans. Sentences which were closely equivalent for us and so synonymous would be remotely equivalent for this creature and so not synonymous for him, although he could still understand them—as we understand the pair of quantified sentences above.

Now let us apply the intensional transparency thesis to this creature with different logical abilities. Suppose that at some stage in a proof, the creature believes that \( a = b \) and \( c = d \); and suppose that we are belief-attributors with "normal" logical abilities as characterized above, so that for us any sentence \( p \& q \) is closely equivalent and synonymous with \( q \& p \). Then because the clause following "believes that" supposedly is intensionally transparent, the creature believes that \( c = d \) and \( a = b \). But it is clear in this case that we, as attributors, cannot always project our particular logical abilities (as opposed to a general level of competence) onto an agent. Because of this creature’s "abnormal" logical abilities explained above, he may not believe that \( c = d \) and \( a = b \); instead, he may be agnostic regarding whether \( c = d \) and \( a = b \).

What difference can there be between the creature’s believing that \( a = b \) and \( c = d \), and his believing that \( c = d \) and \( a = b \)? For, if the creature acts appropriately for the belief that \( a = b \) and \( c = d \), then he will thereby be acting appropriately for the putative belief that \( c = d \) and \( a = b \). The answer is based on the fact that beliefs (in the "strong sense") must be causally efficacious: if an agent has a particular belief, then he must not only act appropriately for that belief, but he must so act because of that belief; in particular, the belief must be part of the agent’s reason for undertaking the action, one of the premises in the practical reasoning which results in the action. We must distinguish between an agent’s merely acting as if he had a particular belief, and his actually having that belief. While the logically deviant creature will act appropriately for the putative belief that \( c = d \) and \( a = b \), as well as for the belief that \( a = b \) and \( c = d \), only the latter will be his reason for these actions—that is, he will only use the belief that \( a = b \) and \( c = d \) as a basis for selecting desirable actions.

Given the creature’s peculiar logical abilities, this situation is no

---

9Davidson discusses this notion in Davidson (1976) and Davidson (1963).
more ruled out than the following case: a normal human being acts appropriately for the supposed belief that for the integers in a particular set, some $x$ in the set is such that for every $y$ in the set, if $x$ is prime then $y$ is odd; and he also acts appropriately for the equivalent belief that if every integer in the set is prime then every one of them is odd—but only the latter belief is the human’s reason for those actions.\textsuperscript{10} Whatever the complexities of the question of what constitutes a reason for an action, it is clear in this case that the human only has the second belief.\textsuperscript{11} The former case is similar; so, correspondingly, the creature there only believes that $a = b$ and $c = d$. If the similarity of the two cases is rejected, then the question must be answered: \textit{why} does the former case have its special status?

Thus, this example shows that the intensional transparency thesis is incorrect: intersubstitution within the scope of a belief clause of expressions which are synonymous, in particular, closely equivalent, for the belief-attributor does not always preserve truth. Such intersubstitution preserves truth \textit{only} on the empirical assumption that:

Expressions which are synonymous for the attributor correspond, in a correct translation, to expressions which are synonymous for the agent.

(The converse is not required.) This assumption is often an acceptable psychological generalization, but the above example shows how it can be violated. If we restrict this assumption to synonyms which are close equivalents, the assumption is a special case of the assumption that inferences which are obvious for a translator are obvious for the subject—i.e., universality of the “lower” portion of the translator’s feasibility ordering. The latter assumption is needed by Quine in his account of translation of logical laws (“Fair translation preserves logical laws”, or better, fair translation preserves obvious logical laws),\textsuperscript{12} and it is unacceptably egocentric even as an approximation for practical purposes, for instance, when teaching logic to a novice. The original argument here against the possibility of a logically deviant creature must beg the question by assuming that there can be no such creature. Granting the limitations of the ‘believes ‘_____’ construction, the ‘believes that _____’ construction must be carefully limited in application when the above assumption regarding synonymies does not hold.

\textsuperscript{10}The sentences expressing the two putative beliefs are of the form, respectively, ‘$(\exists x)(y)(Fx \rightarrow Gy)$’ and ‘$(x)Fx \rightarrow (x)Gx$’.

\textsuperscript{11}There are similar cases involving valid sentences.

\textsuperscript{12}For example, in Ch. 2 of Quine (1960). I discuss Quine’s account in Ch. 5 of Cherniak (1977).
The contingent status of the assumption regarding synonymies must be emphasized; the fact that \( p' \) and \( q' \) are not synonymous for a subject does not by itself rule out their translation respectively as \( p \) and \( q \), where \( p \) and \( q \) are synonymous for the translator. 'Fair translation always preserves all synonymies' is not an acceptable maxim. This is a consequence of a holistic account of meaning of the type mentioned at the end of the last section: The meaning of the expressions \( p' \) and \( q' \) is determined not just by the single ('analytic') equivalence \( p' \leftrightarrow q' \) (or the corresponding inferences), but by the entire range of accepted sentences and inferences in which each expression occurs. In some cases we can establish that the correct translations of \( p' \) and \( q' \) are, respectively, \( p \) and \( q \), if according to that translation \( p' \) and \( q' \) are involved in enough other laws and inferences which we accept besides \( p \leftrightarrow q \).

As we saw earlier, there are some correct empirical generalizations regarding what equivalences are obvious or easy for normal "naive" human beings under certain conditions. But these generalizations have the same status as the earlier assertion about short-term memory capacity. The most we can say is that certain inferences happen to be more intuitive or natural than others for normal humans. We cannot conclude that any rational creature or agent must think in these ways. Generalizations about which are the obvious equivalences are contingent cognitive psychology; a deviant human, or merely one with training in formal logic, can violate these generalizations and still be a logically competent agent. Consequently, a theory of the meaning of an agent's assertions and a theory of what his beliefs are will not together be self-sufficient, contrary to, for instance, Quine (1960); we cannot attribute beliefs and meanings without a third theory—of the agent's cognitive psychology, of how he represents and processes information.

7. Nonverbal Beliefs. One last question is, could there be a feasibility ordering for a creature's inferences from his nonverbal beliefs which differed radically from our own—for example, could commutativity of conjunction fail for a dog's beliefs? In general, the feasibility ordering for a creature's nonverbal beliefs seems to a considerable extent to be a projection of the ordering for the attributor's verbal beliefs.\(^{13}\)

\(^{13}\)This is contrary to the assertion of Bennett (1976, p. 116), "we cannot have grounds for crediting a languageless creature with moderate logical acumen"; in our terminology, only a null or universal feasibility ordering supposedly can apply to such a creature. Briefly, Bennett's conclusion for the type of example we are now considering seems to be based on, among other things, mistakenly taking for granted (pp. 116–117) that the feasibility ordering for a "normal" human's verbal beliefs will be the ordering for the verbal beliefs of any creature.
For, in this situation, there is no unique fact of the matter regarding what is the correct feasibility ordering and what is the correct formulation of the beliefs to assign to the agent—the two are to some extent interconvertible. The choice of a feasibility ordering here is then a matter of which one is not gratuitously complex; but this is a subjective matter, determined by what is familiar to the attributor. So, for a verbal attributor with a radically different feasibility ordering from our own, the rational choice of an ordering to assign to a nonverbal creature (or to our own nonverbal beliefs) would generally not be our own.

Let us consider, for example, a Martian attributor with a feasibility ordering for his verbal beliefs such that for him the inference of $q \land p$ from $p \land q$ was as difficult or unintuitive as the inference of $-(p \land \neg p) \rightarrow (p \land q)$ from $p \land q$ is for us. Suppose that he is informed (and accepts) that Fido believes that it is dinnertime and a cat is on the porch. The assertion that our feasibility ordering universally applies for inferences from nonverbal beliefs will be false in this case. For, this creature will not conclude that Fido believes that a cat is on the porch and it is dinnertime. And it would be as unnatural and gratuitous for him to do so, as would be our concluding from the same original claim that: Fido believes that if it is not the case that it’s dinnertime and it is not dinnertime, then it is dinnertime and a cat is on the porch. Thus, it is a fundamental but contingent fact of our psychology that our own feasibility ordering seems to us to be the “correct” one to apply to inferences involving nonverbal beliefs.

* * *

Our general conclusion, then, is that the holistic interdependence of beliefs and desires and meanings emphasized by Quine and Davidson in fact extends to another domain. Beliefs, desires, and meanings cannot be determined independently of at least a tacit theory of another type: a theory of the agent’s cognitive psychology, of how he thinks. We have been concerned in particular with the theory of the deductive reasoning abilities of the agent, the theory of feasible inferences. A feasibility theory is an indispensable element of a predictive cognitive theory. We have found that there seem to be remarkably few a priori constraints on a rational agent’s deductive abilities: An agent must only be an adequate logician; beyond generalizations about human psychology, there seems to be no transcendentally “right” kind of logician.
REFERENCES