

DEFAULT REASONING: JUMPING TO CONCLUSIONS AND KNOWING WHEN TO THINK TWICE

BY

KENT BACH

Look before you leap. — Proverb

He who hesitates is lost. — Another proverb

PHILOSOPHERS view reasoning as a logical process. Unlike speculating or imagining, reasoning is thinking in the pattern of an argument. Reasoning, good or bad, involves proceeding from premises to conclusion, and these steps comprise the argument being followed. How good the reasoning is depends on the quality of the argument, as determined by the relations, formal or otherwise, that hold among the steps of the argument. On this view to be rational is to be logical. Logic is where to look for an account of good reasoning since, after all, logic is the theory of validity. Ideally, then, reasoning should explicitly follow the pattern of a valid¹ argument.

Philosophers also generally suppose, and I unskeptically agree, that ordinary people are capable of good reasoning and often indulge in it. And they suppose further that when reasoning falls short of this standard, that is due to substandard performance, not to outright incompetence. People may make mistakes, cut corners, get distracted, or just not care, but presumably anyone of normal intelligence and decent education is capable of good reasoning. Reasoning does not have to be as good as possible to be good, but supposedly the closer it is to the ideal (whatever "closer" means here), the better it is. Thus, so the common conception goes, the more we strive to reason ideally, the better we will reason.

I believe that this conception is mistaken. Reasoning is not in general a matter of proceeding explicitly from premises to conclusion.² How good it is is not a matter of how closely it approximates ideal reasoning, at least if that means explicitly following a valid pattern of argument. Indeed, on the conception of reasoning I will develop, it would not even be good cognitive policy generally to try to reason ideally. Even if there were a logic of ideal reasoning, trying to

follow it would still be unreasonable. That is partly due to our cognitive limitations but is due ultimately to a general feature of the human condition.

1. What Is Default Reasoning?

I once asked an eleven-year-old girl called Polly, "what problem are people faced with at every waking moment?" She immediately replied, "the problem of what to do next."³ Polly was right: we *are* always faced with the problem of what to do next, certainly if doing includes mental actions, such as directing one's attention or making a decision. Of course, the problem of what to do next breaks down into an endless succession of little problems. As agents, we generally do not consciously attend to these little problems but concentrate on those worthy of deliberation. As theorists, we focus on deliberate thought and action, often pretending that agents are faced with one clearcut problem at a time, are presented with well-defined alternatives for consideration, and can make a judgment or decision based on an explicit evaluation of them. But all this takes time, and at any given moment the problem of what to do next must be solved immediately. Usually we solve it without much trouble, thanks to our extensive knowledge of what to do when (as contained in recipes, rules of thumb, habits, and skills).⁴

Any plausible conception of human reasoning must reckon with the real-life constraints, internal and external, under which we generally must make our judgments and decisions. Most of that reasoning is not rigorous and detailed but casual and streamlined, relying on implicit assumptions, rules of thumb, and other inferential shortcuts. If, as I will assume, commonsense reasoning is commonly reasonable, then finding out how it works could well tell us something about what makes it reasonable. Moreover, no *a priori* normative theory of reasoning could be correct if it required cognitive processes to work quite differently from how they actually work. Accordingly, a theory of good reasoning that is meant to apply to actual human reasoning must take the nature of these processes into account.⁵

Philosophers may suppose that the more explicit a piece of reasoning, the better it is, and that inexplicit reasoning can always be improved upon by being made more explicit. Yet most of our reasoning, including much that gives us knowledge, is largely inexplicit.⁶ People in AI recognize this, as is evident from their extensive work on so-called *default reasoning*. Philosophers do discuss related notions, such as rules of thumb and *ceteris paribus* rules, and in ethics and epistemology they recognize that the reasons on which we base our judgments are often defeasible. Yet they don't consider the implications of such rules and reasons for the nature of reasoning itself.

I suggest that we think of default reasoning as "inference to the first unchallenged alternative." By this I mean the first option that comes to mind without being immediately followed by the thought of a reason against it or of an alternative to it. I don't mean that we check for the occurrence of such a thought, for what is crucial is not our noticing the thought but its very occurrence. In this way, we jump to conclusions except when we look before we leap.

The simplest kind of default reasoning occurs when, as a question comes up, we believe the first thing that comes into our heads. "Reasoning" might seem too strong a word for this pervasive phenomenon, but I believe it deserves the

name. Whatever we call the processes that lead to them, snap judgments need not be arbitrary or capricious but are often justified and can yield knowledge in the full sense of the word. Later I will argue that the process that leads to them is rule-governed, involves implicit assumptions, and can lead to justified beliefs insofar as it is reliable. Reliability here involves knowing when not to jump to conclusions but instead to think twice.

I will make analogous points about practical reasoning and snap decisions. Snap decisions are reasonable provided we know when not to make them. Most of the time we do something when the thought of doing it occurs and no thought to the contrary follows. In other words, we do it unless the psychological path from thinking of doing it to doing it is broken by some intervening thought.⁷ I would go so far as to suggest that at every waking moment what we do next is the first thing that occurs to us to do, provided the thought of doing it is not immediately overridden by some further thought. In practical terms, for something to seem good enough to do at a given moment consists simply in the unoverridden thought to do it.

Notice that I am taking conclusions of factual⁸ reasoning to be judgments and conclusions of practical reasoning to be decisions. The usual view seems to be that factual conclusions are beliefs and practical ones are intentions, but this seems wrong in that beliefs and intentions are not events but states. Conclusions must be events, since they are outputs of real-time cognitive processes, and judgments and decisions fit the bill. Of course judgments normally result in beliefs and decisions in intentions.⁹

The process leading to a snap judgment or decision is called *default reasoning* because of how its conclusion is reached. Its conclusion is arrived at by default in the sense that the reasoning is based on some generalization or stereotype which is overridden only if there occurs the thought of an alternative or of a reason to the contrary. The stereotypical assumption is like the default value assigned to a variable in a computer program. When a value needs to be assigned, the default value is assigned automatically if no alternative is provided, and the program runs from there.

Reasoning leading to a snap judgment or decision is not the only kind of default reasoning. More elaborate reasoning, containing a number of steps, can still be default reasoning, for there can be an implicit assumption at any step along the way. We make such an assumption whenever we reason in a way that is sensitive to it: drawing inferences consistent with it and not drawing ones inconsistent with it. Ordinarily we do not question such an assumption unless there occurs to us some reason to do so. We rely on our ability to detect or to think of reasons, when worth considering, for challenging our assumptions. For example, we often apply generalizations automatically and yet, relying on our ability to detect exceptions, we often know when not to apply them. These abilities can become highly refined, as with experts like detectives and doctors (as modeled by expert systems in AI), but we are all experts about many aspects of the world around us.

When our reasoning to a conclusion is sufficiently complex, we do not survey the entire argument for validity. We go more or less step by step, and as we proceed, we assume that if each step follows from what precedes, nothing has gone wrong. That is not always so, for an implausible conclusion along the way may lead us to question some previous step (either a premise or a bit of reasoning).

An intermediate conclusion will seem implausible if it conflicts with other beliefs. Of course there is no guarantee that we will detect every such conflict, but we implicitly assume that when there is one, we will detect it and go back over our reasoning. Here we rely on our ability to detect such conflicts. Even if our lines of reasoning were always perspicuous, so that we could view them as a whole, there would still be points at which we do not actually check for validity but simply "go along" with the reasoning at that point. We just "see" that the next step follows.¹⁰ Reasoning does not have to be evaluated in every evaluable respect in order to lead to justified belief. It can include elements that are not explicitly evaluated, not to mention implicit assumptions that would become explicit if the step in question were explicitly evaluated. Their implicit "evaluation" consists simply in their not being questioned. Such an evaluation is reliable insofar as one is reliable at detecting good reasons for questioning steps in one's reasoning.

In sum, default reasoning is reasoning that contains at least one defeasible step, and what that is can be described intuitively as follows. When you take such a step you do not think, "Everything's OK, so I'll take this step." Rather, you just take it unless you think something might be not be OK. The quality of your reasoning at this step depends on your ability to know when not to take steps of that sort.

II. Taking Things for Granted

My interest in default reasoning arose from a psychological issue raised by Gilbert Harman's ingenious approach to the so-called Gettier problem in the theory of knowledge. Edmund Gettier (1963) showed the traditional definition of knowledge, as justified true belief, to be insufficient. He discovered that two beliefs, both true, can be justified in just the same way even though one qualifies as knowledge while the other does not. Here is a simple, well-known example (Goldman 1976, 772–3). You are driving along the countryside and spot what looks like a barn. You judge it to be a barn and in fact it is. Your belief is justified and if the situation is normal, you know it to be a barn (only a skeptic would deny this, but the Gettier problem does not concern skepticism). However, suppose there happen to be some realistic fake barn façades in the vicinity. In this extraordinary situation you do not know that what you see is a barn, even though your true belief (you happen to be looking at a real barn) is justified in a way that normally suffices for knowledge. The Gettier problem is to explain the difference between a justified true belief that qualifies as knowledge and its "gettiered" counterpart.

Unlike other attempted solutions to the problem,¹¹ Harman's approach in *Thought*¹² is conceptually simple and elegant. He suggests that the above difference can be explained by principle P, "Reasoning that essentially involves false conclusions, intermediate or final, cannot give one knowledge" (47). In the gettiered barn example, the culprit is the proposition that there are no barn façades around. In the normal case, of course, that proposition is true. Now Harman holds (29–31), and I agree, that for a belief to be justified by a piece of reasoning, the belief must result from the reasoning. If so, principle P can discriminate between the two beliefs only if the reasoning leading to each actually contains the proposition that there are no barn façades around, for this is the proposition

true in the normal situation and false in the other. Harman recognizes the obvious objection that the reasoning does not seem to contain such a proposition but, he insists, being aware of a premise is not necessary for reasoning from it (21). Indeed, it is Harman's strategy "to turn skepticism on its head and use intuitive judgments about when people know things to discover when reasoning occurs and what its principles are" (112). Recognizing that a belief must be justified to qualify as knowledge and insisting that its justification depends on the reasoning that actually led to it, Harman proposes a novel form of psychologism in which, at least as a first approximation, "the valid principles of inference are those principles in accordance with which the mind works" (18).¹³

As mentioned, Harman is unconcerned that in Gettier examples the person seems not even to consider the proposition (true in the normal case, false in the gettiered case) that P says he must believe. Perhaps the proposition does occur unconsciously, but only Harman's strategy suggests this. Considerations of psychological plausibility¹⁴ suggest otherwise, and surely they can't be overridden by the appeal of a neat solution to the Gettier problem. Harman's solution may require supposing that in the barn example the person's reasoning contains the proposition that there are no barn façades in the vicinity, but this supposition is not independently plausible. As I will suggest, however, Harman's strategy can be made plausible if modified to allow for propositions that do not explicitly occur in a process of reasoning, not even unconsciously, but are merely implicitly assumed. They are taken for granted.

Consider one of Harman's own Gettier examples. A person seems to be seeing a candle and justifiably believes, correctly in fact, that there is a candle in front of him. Normally he knows this but in the gettiered case he does not. For although there is candle in front of him, he is unaware (and has no reason to suspect) that what he is seeing is really the reflection in a mirror of a candle off to the side. In inferring that there is a candle in front of him, he does not seem to be explicitly supposing that what he is seeing is not the reflection of a candle. No wonder Michael Williams (1978) has argued that there is no evidence to warrant ascribing any reasoning that includes this supposition.¹⁵

Can Harman's strategy be defended against this charge of psychological implausibility? I believe it can be, with the help of a crucial psychological distinction and an appeal to a rule of inference that requires this distinction. In defending Harman's strategy, I will be following the spirit of his psychologism, according to which, at least as a first approximation, rules governing how the mind works are valid rules of inference.

THE DISTINCTION

According to Harman the identity of a process of reasoning is determined by the abstract structure of inference it instantiates (48). Moreover, because he sees no reason to expect systematic type–type correlations between neurophysiological and psychological states, even for a single person. Harman doubts that any amount of neurological information could provide a reliable basis for ascribing reasoning (20). His psychologistic strategy is supposed to provide such a basis, but unfortunately he does not tell us much about the nature of the relationship between a process of reasoning and the abstract inference pattern it

instantiates. In particular, I wonder, must this process contain elements (whether specified psychologically or neurophysiologically) corresponding to all of the steps in the inference pattern?

Offhand, you might suppose that it must. Surely there couldn't be two pieces of reasoning that contain the same elements but do not instantiate the same inference pattern. Nevertheless, we can draw a certain psychological distinction such that two processes of reasoning containing the same elements *can* instantiate two different inference patterns, one containing a step the other lacks. A piece of reasoning fully *realizes* an abstract pattern of inference if it contains psychologically real elements corresponding to all the steps of that pattern. It merely *instantiates* that pattern if there is some step that is not explicitly included but is merely implicitly assumed.¹⁶ To say that a person implicitly assumes a certain proposition in his reasoning means that the reasoning would not occur unless he believed that proposition. Thus, if he explicitly reasons from *p* to *q*, he is implicitly reasoning from *p* and *r* to *q* if he would not have explicitly reasoned from *p* to *q* had there occurred the thought of not-*r* or of a reason against *r*.¹⁷ In this way his explicit reasoning, though direct from *p* to *q*, is not based on the belief that if *p* then *q*, which he does not hold anyway.

Thus we can make sense of the idea of an implicit assumption as one that underlies a process of reasoning without actually being included in it, and we can thereby explain how two processes of reasoning explicitly containing the same elements can instantiate two different inference patterns, one containing a step (the implicit assumption) the other lacks.¹⁸ This idea is illustrated by routine kinds of belief formation, as occurs in perception, communication, and action whenever beliefs are formed not deliberately but spontaneously (that does not mean arbitrarily). If you introspect what happens, it seems as though you go directly from an experience to a belief: you see that there is a candle ahead, you recognize that a speaker is asking you something, and you expect a match you just struck to light, but you don't notice (if nothing seems out of the ordinary) any intermediate steps in your reasoning. Yet it might seem that if such beliefs are justified, there must be such steps. After all, seeming to see a candle does not justify believing a candle is there. Harman thinks the belief is justified only if you suppose that perceptual conditions are fairly normal, that your vision is in good order, etc. So if such suppositions are included in your inference, which presumably is justified, they must be unconscious. Harman (1973, 21) has no qualms about that; indeed, in regard to these ordinary, routine cases, he cheerfully acknowledges that his strategy requires ascribing unconscious reasoning to people. He is right to insist that not noticing intermediate steps does not mean they don't occur. Perhaps the sheer speed with which you go through them explains your inability to notice them, just as in producing and comprehending sentences you don't notice yourself following grammatical rules. However, there is another possibility: perhaps you just skip the intermediate steps and simply jump to conclusion—and do it justifiably.

How could this be? Take the perceptual example.¹⁹ There appears to be a candle before me, and I infer that there is one. I seem to do nothing in between. Yet Harman maintains that because of the gettiered version of the same example, "the perceiver's reasoning requires the conclusion that whatever accounts for there normally being a candle in front of me, when it looks as if there is one,

explains why there is a candle in front of me now, when it looks as if there is one." (Harman 1978, 436) Ascribing such an intermediate step is required if principle P is to explain why the gettiered belief does not qualify as knowledge and also, in Harman's view, to account for the belief's being justified. Still, even disregarding introspection, it is psychologically implausible to suppose that one's reasoning contains such a step. The belief that there is a candle is but one of countless beliefs that I form continuously as I contemplate my surroundings or navigate about them, and it seems highly unlikely that for each and every object I take to be before me, I draw a distinct intermediate conclusion like the one Harman suggests for the case of the candle. That seems not only implausible but highly inefficient. Much more plausible to ascribe and efficient to use would be a generalization like this: ordinarily things are as they seem, because they seem as they do because of the way they are. This generalization could be used over and over, as countless perceptual judgments are made. It would function as an intermediate step each time one infers the presence of something, but a new intermediate step would not be needed for each new inference. There is no reason to suppose, Harman's strategy notwithstanding, that this same step explicitly occurs over and over, as each succeeding judgment is made. And yet such truncated reasoning, though lacking an element corresponding to the intermediate step, could be both explicable and justified, provided that the intermediate step is implicit in it.

THE TAKE-FOR-GRANTED RULE

Because there appears to be a candle before me, I directly infer that there is a candle before me. I do it because I have long since learned to draw such conclusions directly. In general, whenever it appears to me that *p*, I directly infer that *p*. Only if it occurs to me that something other than the fact that *p* might explain its appearing that *p* does it occur to me to rule out such a possibility before inferring that *p*. But most of the time nothing of the sort occurs; being neither a paranoiac nor a compulsive epistemologist, I infer that *p* directly. Yet I don't assume that if it seems to me that *p*, then *p*. Rather, I seem to reason according to something like the following rule, which I call the *take-for-granted rule*, because in following it I take for granted (implicitly assume) something needed to justify what I directly infer.

(TFG) If it seems to me that *p*, then infer that *p*,
provided no reason to the contrary occurs to me.

However, in directly inferring that *p*, I do not *think* that no reason against *p* is occurring to me. Rather, whenever no such reason comes to mind, I simply infer that *p*.

I am suggesting, then, that my reasoning is governed by TFG even though I do not have to think that the condition it specifies obtains. That is, my reasoning does not contain the thought that no reason against *p* is occurring to me. And yet it instantiates an inference structure that does include that step. It does so because it is governed by TFG rather than by the simpler but dubious rule that if it appears to me that *p*, then infer that *p*. In this way, reasoning governed by TFG can instantiate an inference pattern without fully realizing that pattern.²⁰

DEFAULT REASONING AND JUSTIFIED BELIEF

How can such abbreviated reasoning yield justified beliefs and, if the belief is true and ungettiered, give one knowledge? For example, in the ungettiered case of the candle, how can I know that there is a candle in front of me simply by inferring this from how things appear? I couldn't know this if I believed, for example, that objects always are as they appear, that is, if I followed the preposterous appearance-is-reality rule (AIR): If it appears to me that *p*, then infer that *p*. If I did follow AIR, then when I inferred that something is as it appears my reasoning would instantiate a simpler (and obviously invalid) inference pattern than the one it does in fact. Fortunately, my reasoning follows a different rule, TFG, and generally instantiates a valid inference pattern, one that is validated, I suggest, by what we might call the *take-for-granted principle*:

(TFGP) Its appearing to one that *p* justifies directly inferring that *p* provided that

- (a) it does not occur to one that the situation might be out of the ordinary, and
- (b) if the situation were out of the ordinary, it probably would occur to one that the situation might be out of the ordinary.

(The force of "ordinary" here is to exclude sources of illusion, distortion, and hallucination.) When TFGP applies, I am justified in taking for granted that the situation is ordinary, unless it occurs to me that perhaps my perception is being affected abnormally, say by bad lighting or by devious psychologists. Thus, as clause (b) provides, TFGP licenses my implicit use of TFG to the extent that I am able to detect abnormal circumstances. I must be pretty good at knowing when not to infer that things are as they seem in order to be justified, when the situation is normal, in supposing that things are as they seem. If I were insensitive to abnormal situations, I would directly infer that *p* even when I should not.²¹

TFGP might seem objectionable in that the reasoning it is supposed to license is conditional on the nonoccurrence of a certain thought. How could the reasoning be rendered invalid merely by the occurrence of that thought? Yet this is what TFGP says! It is a valid principle because, as a matter of psychological fact, I, like most anyone else, am equipped to detect evidence that a given perceptual situation is out of the ordinary. I cannot do this infallibly, of course—undetected illusions do occur—but I can do it reliably. Whenever I directly infer that things are as they appear, I rely on my ability to detect abnormalities. If I am justified in so doing, then in true psychologistic fashion TFGP not only captures a fact about how my mind works but also justifies the conclusions I draw directly from how things appear.

As for TFG, the rule whose use TFGP licenses, it has the remarkable feature that it cannot be explicitly followed.²² By this I mean not just that it is followed unconsciously but, in particular, that following it does not involve checking to see if the thought of any reason against *p* has occurred. If it did, then the question whether TFG applies to the situation would occur to me, in which case my reasoning would not be direct. I wouldn't be following TFG, and my reasoning would not be justified, at least not by TFGP. In any case, ordinarily the question whether the situation is normal does not occur to me. That is, whenever it does

not occur to me that the situation is out of the ordinary, it also does not occur to me that the situation is ordinary. And when no such thought occurs, generally the situation *is* ordinary.

Ordinary perceptual judgments are but one example of how we commonly and reliably jump to conclusions (except when we look before we leap). We seem to be generally right in our snap judgments not only about the kinds and qualities of things we perceive around us but also about vivid past experiences, names of familiar persons and places, about spellings of words, and about what people mean when they speak to us (Bach and Harnish 1979). Thus, just as I assume that things are as they look, unless it occurs to me that my vision is being affected abnormally, so in the case of recall, as of somebody's name or the spelling of a certain word, I take for granted that the first thought that comes to mind is the right one—unless some reason against that or some alternative comes to mind. Thus it seems that just as in perception we follow TFG, so in other areas we follow counterparts of TFG.

Jumping to conclusions is efficient, but why should it be reliable? Offhand, jumping to conclusions would seem to gain speed only at the risk of error. There seems to be a tension between efficiency and reliability, but don't forget, as Mill observed, that drawing inferences is "the only occupation in which the mind never ceases to engaged." We can't avoid trading off possible error for speed, for there are always more inferences to be made. If we didn't generally jump to conclusions, we would make hardly any of the inferences we needed to make. But if we are generally right in the conclusions we jump to, surely this is no monumental coincidence. Somehow our inferences must take relevant information into account without getting bogged down in irrelevancies. The tension between reliability and efficiency is resolved by the fact that alternatives can be effectively and legitimately ruled out without even being considered. This means that we must be equipped with backup processes for detecting when the condition presupposed by the rule is not met. I have said nothing about how these processes work, how they are organized, or how they are implemented, and obviously all this needs explanation. But whatever the psychological account of their workings, there must be some explanation for them. Otherwise we could not rely on their reliability!

III. Snap Decisions

Decision theorists and philosophers of practical reasoning often make a number of idealizing assumptions: (1) that a given problem is explicitly posed to a person, (2) that a set of possible solutions is provided, (3) that relevant information is available and is presented as relevant, (4) that there is no time limit on deciding, and (5) that the decision cost is negligible. Sometimes one or two of these assumptions is relaxed, as in the study of decisionmaking under uncertainty, but by and large they are implicit in various theoretical conceptions of practical rationality. In reality, most problems do not come neatly defined and cannot be dealt with at one's leisure. Besides, there is always the problem of what to do next. Ordinarily we are not aware of this brute (but generally not brutal) fact of life, thanks to the further fact that we usually have a reasonably good idea of what to do next and do it immediately. Consider acts like turning a doorknob,

putting on the brakes, and writing your name. There are many types of action such that when we consider performing one, we generally decide to perform it immediately and don't have to think how. This is a fact for any account of practical reasoning to reckon with. It can, I will suggest, by appealing to practical analogues of the TFG rule and the TFGP.

To appreciate the need for such a rule, consider for a moment the following misguided question about what is involved in acting immediately (without deliberating). Before deciding what to do, do you first decide to do immediately whatever you thereupon decide to do? That is, do you make a choice between acting immediately, doing the first thing that comes into your head, or deliberating about what to do? This absurd question suggests that when you act without hesitating, you do so as the result of deliberation. Worse, it assumes that when you do decide to act immediately rather than to deliberate, you make two separate decisions, first to act immediately and then what to do. Surely this picture is all wrong. Rather, you do what occurs to you to do unless it immediately occurs to you not to, or at least the thought of an alternative or the thought of a reason against doing it. In any case, the psychological path from thinking of it to doing it is broken by some intervening thought. I am suggesting, then, that at every moment we next do the first thing that occurs to us, provided this thought is not then overridden by a further thought. Of course, there are many things that we do as the result of deciding to do them, but most of what we do we do because we don't decide not to. That is, usually we do something because it occurs to us to do so and it does not next occur to us not to.²³ On my word processor I must punch the EXECUTE key in order to get it to do something (that is, to carry out a command I've entered), but I don't have to do any such thing to get myself to do something. For example, I feel cold, so I get up and close the window. It doesn't occur to me not to. This might have occurred to me, but it doesn't. So I act. The same thing happened to William James when staying in bed one cold morning. Every so often, he thought about getting up. Eventually he did. Evidently, every previous time that he thought about getting up he thought also of a reason not to and didn't. The last time no such thought occurred.

LEAPING WITHOUT LOOKING

There is a counterpart to TFG that governs direct practical reasoning. Consider that in thinking about what to do, we would never be able to make a decision if we first had to rule out every consideration contrary to what we contemplated doing and had to decide against every alternative course of action. Fortunately for most of us most of the time (but not often enough for some), we don't consider what's not worth considering, not even to determine that it's not worth considering. We seem to follow the *not-worth-considering rule*:

(NWC) If it occurs to me to do A, do A unless there occurs to me the thought of a reason to the contrary or of an alternative to A.

By following NWC we make snap decisions, not bothering to consider possible reasons against them or alternatives to them. Of course, such considerations often are worth addressing, and it would be unfortunate, to say the least, if we always did the first thing that came into our heads. Fortunately, we often do recognize when what is ordinarily not worth considering is so under the special circum-

stances at hand. For example, if it occurs to me to drink some milk, usually I do so without further ado. I don't check to see if it is bad. Nevertheless, if I have been away for a week I think to sniff it before drinking it; if it smells bad, I don't drink it. Insofar as we are good at knowing when to consider what is ordinarily not worth considering, in normal circumstances we are justified in not considering such things. This is captured by the practical counterpart of TFGP, the *not-worth-considering principle*:

(NWCP) If it occurs to one to do A, one is justified in directly deciding to do A, provided that

- (a) there occurs to one no thought of a reason to the contrary or of an alternative to A, and
- (b) such a thought probably would occur if it should.

As we saw with TFGP, NWCP is valid only to the extent that one is good at knowing what is worth considering in decision contexts. Of course, what this comes to depends on what counts as worth considering. That is a matter of one's interests and values, and even one's personality. A carefree person has much lower standards than a worrywart. What is worth considering also depends on what the environment is like, on how friendly or hostile it is.

We cannot spend the bulk of our time sitting in cognitive cocoons and contemplating our options. Many of our actions result from deliberation, but many more do not. That is why I have developed a causal theory of action (Bach 1978) that deals not only with deliberate actions but also with those that are not deliberate and not even intended.²⁴ A nondeliberate action is the immediate result of the thought of performing it. And most actions, including the minimal action components of deliberate actions, are like this. The thought of a certain action occurs to us and we just perform it—unless there occurs a reason against it or the thought of an alternative. For example, when you speak, most of the words you utter just "come out" (not that there isn't an explanation for which ones come out). It's exceptional when you pause to think of just the right word before coming out with one. Similarly, when you look around and move about a safe, familiar environment or do such routine things as putting on your shoes or brushing your teeth, for the most part you do not think before you act. Fortunately, though from an evolutionary point of view not surprisingly (given that the general features of our surroundings are as they are), we are so equipped psychologically and biologically that most moments of our lives we can safely think and act undeliberately.

LOOKING BEFORE LEAPING

Even when an action is deliberate, the reasoning that leads to it is not likely to be fully explicit. It may follow NWC every step of the way. If you are trying to decide what to do some bright Sunday afternoon, how do you know when to stop thinking about it? At any given point, you might be leaning toward one particular option, say going to the beach, but you have not yet settled on that option. Nor have you set a deadline yet for your decision (you might, before your opportunities slip away). The decisionmaking process is open-ended, in the sense that more can always go into it: invention of new options, deeper search into the consequences of the old options, reevaluation of your assumptions or preferences.

Presumably the length and the depth of the decisionmaking process depends on the assumed importance of the decision, together with situational time limits, but even so, there is generally no precise point at which the process must end. Yet it often does end, and with a decisive result.

One suggestion how this happens is that there is an ongoing measure of the decision cost, which is compared to the value of the decision itself. When the cost gets too great relative to the value of the decision (which itself might diminish as opportunities get cut off and options foreclosed), a decision is made: the option highest on the list gets selected. That seems plausible enough, but who knows if it really happens? Here's another suggestion. The decision is made when, after exactly one option is at the top of the list, the decisionmaking process stops. It stops when no further considerations occur to you: no further options, no reconsideration of previously considered options, no reasons for not selecting the currently preferred option. In other words, you decide on the preferred option when nothing against it occurs to you. At that point, by following NWC you select it by default.

IV. *The Dynamic Roles of Beliefs and Intentions*

Truncated reasoning that follows default rules like TFG and NWC depends on the nonoccurrence of certain specified thoughts. Such reasoning is justified only to the extent that the relevant thought would occur if it should occur. Somehow, we are so constituted that commonly our antecedent attitudes are automatically brought to bear upon our reasoning when they are relevant to it. However, so far as I know, no account has yet been put forth of how beliefs and intentions get called up to the service of our reasoning. I have no such account to offer, but I will make a few elementary points about belief and intention that seem relevant to such an account. Although I maintained at the outset that judgments and decisions, not beliefs and intentions, are the conclusions of factual (theoretical) and practical reasoning, beliefs normally result from judgments and intentions from decisions. When activated they play roles in our reasoning. Philosophers of mind may speak of the causal or functional roles of these states, but they seem to mean static, conceptual roles. Here I will describe the dynamic roles of beliefs and intentions and make some comparisons between them along the way.

To believe something is to be prepared, when the thought of it occurs, to assume it and reason from it without reconsidering it (unless, of course, some new consideration occurs to you). Thus if you believe that *p* and the thought of it occurs, you will take it that *p* without further ado.²⁵ Because you already believe it, the question of whether or not *p* does not arise when the thought of *p* occurs; the question is answered already. Moreover, part of the function of a belief is to come to mind when it is relevant to what is currently at issue. Notice that you do not have to remember that you believe that *p*; you just have to remember that *p*.

Just as having a belief enables you to know what to think without having to reconsider, so having an intention enables you to know what to do, again without having to reconsider. As Stuart Hampshire puts it, "the point of having a firm and fixed intention is that I do not need to think further about what I am to do,

as the point of having a firm belief is that I do not need to trouble myself further to wonder whether a statement is true" (1960, 101). So if I intend to do A (at a certain time or on a certain occasion), then when I notice that the time has come, generally I will do A without deliberating or rehearsing my previous deliberation. And, as with believing, I do not have to remember that I intend to do A in order to do it automatically when the time comes; I just have to remember to do A. Indeed, if I am not anticipating the time or occasion of action already, the intention to do A tends to prompt me to notice that the time has come.²⁶ When I notice, because A is what I intend to do then, I do not reconsider whether or not to do A—unless a (presumably) new consideration against doing A occurs to me.

Believing that *p* does not mean being unwilling to reconsider it when the thought of it occurs. Unless your belief that *p* is not just strong but downright dogmatic, new considerations relevant to the question of whether or not *p* can lead you to reconsider the question, even if you end up retaining your belief. Intending is similar: intending to do A does not mean being determined to do it no matter what, and new considerations that come to light can lead you to change your mind. This could be the result of a change of heart (desires, preferences) or be due to beliefs acquired since the intention was formed.

Beliefs and intentions play similar roles in reasoning. Believing that *p* enables you to make inferences concerning matters (factual or practical) to which the question of whether or not *p* is relevant, often without consciously having to think that *p*. Of course there is no guarantee that you will rely on it in reasoning about every matter to which it is relevant. And even when you do rely on it, there is no guarantee that it will survive the reasoning in which it begins as a premise, for rather than accept what it leads to you might abandon it instead. Like beliefs, intentions play a role in reasoning. For if you intend to do A, your intention could well be relevant to your practical reasoning about other actions (also, since intending involves believing (Harman 1976, 432 ff), your intention could well be relevant to factual reasoning about the future). This practical reasoning might be planning, where your doing A is part of a larger scheme, or it might be deliberation about actions not directly related to A, where what matters is not A itself but the time, place, cost, or consequences of doing A. Reasoning on other practical matters may lead you to give up the intention of doing A.

Finally, there is the question of degree or strength of belief and intention. Beliefs come in a variety of strengths ranging from hesitance to conviction, and a rough measure of a belief's strength is its resistance to being given up in the face of other beliefs. Even though it comes in degrees (or strengths), belief is still all-or-nothing. Of course you can believe that probably *p* (to some qualitative or even quantitative degree), but this is not the same as believing to a certain degree that *p*. Thus degree of belief should not be equated with subjective probability.²⁷ One can take *p* to be much more probable than not-*p* without believing it at all, and even believe some other proposition *q* that one takes to be less probable than *p*. Like beliefs, intentions can be more or less firm, and strength of intention in this sense should not be confused with expected utility or with strength of desire. This is not to say that strength of intention is all that clear-cut. Measuring strength by the likelihood of an intention's being acted on could give one result while measuring it by competitive edge (in reasoning and planning) over other intentions could give another.²⁸

KINDS OF INTENTIONS

In comparing intentions with beliefs, I had in mind simple or, as I prefer to call them, *categorical* intentions. The content of a categorical intention is unrestricted except for specifying when the action is to be performed. That is, a categorical intention is the intention to perform a certain action at a certain time or on some expected occasion. I distinguish categorical intentions from two other kinds, *conditional* and *qualified* intentions.²⁹

A conditional intention is different in content from a categorical one, even though the latter may be tied to some future occasion. An intention is categorical if one expects the occasion on which the action is to be performed to arise either in the natural course of events or as part of one's plans. For example, I categorically intend to pay my rent on the first of the month or to offer a drink to my guest whenever she arrives. Whereas such an intention is of the form 'I intend to do A when C,' a conditional intention is of the form 'I intend to do A if C.' It is a contingency plan. The fulfillment of the specified condition provides not merely the occasion of action but the reason for action. And one might not expect this condition to arise. For example, I conditionally intend to call the fire department in the event of a fire, but I do not expect there to be a fire nor do I have a reason to start one. But if I categorically intend to cook hot dogs when the campfire is going strong and I do not expect the campfire to start of its own accord, I have reason to start it, to incorporate building the fire into my plans.

A qualified intention is the intention to do something *unless* a certain condition arises. For example, believing that it might rain later on, you might intend to play tennis this afternoon unless it rains (and you might have a contingency plan for that, say to play basketball). Now very few intentions are intentions to do something no matter what. So it might seem that almost every intention, if not conditional in the above sense, is a qualified intention. That is, for almost every seemingly categorical intention, there is some condition(s) under which one does not intend to perform the intended action. However, in general this condition is not specified in the content of the intention. For example, you might intend categorically to go to Europe this summer knowing full well that you would not go if you were financially ruined or sent to prison in the meantime, but such conditions as these are not included in the content of your specific intention to go to Europe. You have no reason to think they might obtain, and they have no specific relevance to your plans—they would be relevant to any plan that requires financial and personal freedom to carry out. In contrast, your intention might specifically take your limited budget into account. So, for example, you might intend to go to Europe unless the cost of the trip should exceed \$2000. In the same way your detailed vacation plans might include various sub-intentions, like going to Berlin provided you can book a room in the Hotel Bristol Kempinski (with plans like that your \$2000 trip will not be a long one!).

Earlier I described some dynamic roles of beliefs and intentions generally. Because of the conditions specified in their contents, conditional and qualified intentions play special roles. When it occurs to one that the condition of a conditional intention obtains, one performs the conditionally intended act. When it occurs to one that the condition of a qualified intention obtains, one refrains from performing the intended act. (Of course in both cases new considerations may come to mind and override the intention.)

V. The Frame and Related Problems in AI

Default rules like TFG and NCW are germane to the so-called *frame problem* in AI.³⁰ This is "the problem of maintaining an appropriate informational context, or frame of reference, at each stage during problem-solving [or planning] processes" (Raphael 1976, 147). Whenever we perform an action, form a belief, or consider a hypothetical possibility (in planning or theorizing), there are countless consequences that do not occur in our reasoning but would be obvious, given our entire system of beliefs, if considered. Yet we need to infer them if we are to maintain an accurate view of the world. However, it would be impracticable to infer them all explicitly. Some may require attending to and therefore must be inferred explicitly, but the rest need only be inferred implicitly.³¹

In AI the frame problem is part of the general problem of *relevant inference*, of getting a system to use available information to make relevant inferences without getting bogged down in irrelevant ones. People are pretty good at this. For example, if you have just dressed for a formal dinner and then discover your car isn't running, you realize that if you take your motorcycle, you'll mess up your tuxedo. So you call a cab. Your beliefs also warrant the inference that you are not wearing a football uniform. Yet this thought, not to mention countless others that are equally warranted, does not occur to you. That is fortunate, as there is no end to the number of irrelevant inferences you could make. Of course not everyone is good at making relevant inferences and at not making irrelevant ones. Obsessive-compulsives have much too liberal a notion of what is worth considering, so that almost anything is relevant and things never seem to get done. On the opposite extreme are impulsive people, who fail to consider relevant consequences of their thoughts or deeds.

We can get a system to make relevant inferences only if we can give it some way of identifying what is relevant, but virtually anything that is generally irrelevant to a task of a certain sort can be relevant to it in some contexts. And Marvin Minsky has observed,

Even if we formulate relevancy restrictions, logistic systems have a problem using them. In any logistic system, all the axioms are necessarily "permissive"—they all help to permit new inferences to be drawn. Each added axiom means more theorems; none can disappear. There simply is no direct way to add information to tell such a system about kinds of conclusions that should *not* be drawn! . . . If we try to change this by adding axioms about relevancy, we still produce all the unwanted theorems, plus annoying statements about their irrelevancy. (1974, 125)

Indeed, I should add, most of these statements would not just be annoying but themselves irrelevant (most of them are "no-change" or *frame axioms*). A variety of so-called nonmonotonic logics³² have been proposed to deal with the problem of relevant inference, but none has solved Minsky's problem. Minsky himself has proposed a nonlogistic way around this problem, which I'll sketch after mentioning two more related problems.

Akin to the problem of relevant inference are the *qualification* problem (the problem of what to assume) and the problem of *what to take into account*. In many practical situations, because of the pressure of having to take action we make unchecked assumptions, often acting as if they are true rather than actually accepting them. Among these are assumptions about which information, ideas,

and approaches are worth considering and which are not. We seem to operate with heuristics or rules of thumb, call them what you will, that in relevant ways abbreviate processes of inquiry, inference, and deliberation. Like *ceteris paribus* statements and default rules, they are reliable but not foolproof. Fortunately, when things go wrong, backup systems, invoking troubleshooting heuristics, can come to the rescue, as illustrated by recent work in AI (such as Lenat 1982, 1983 and Aikins 1983). But the general question posed by the above problems remains: Of all the information available to us in memory and in perception, how do we know which information to use (and when) and which inferences to make?

Recall that following TFG or NWC to make snap judgments or decisions requires the nonoccurrence of a certain thought. That is why TFG and NWC cannot be explicitly followed. I don't make my judgment or decision after checking to see that no thought of an alternative or contrary consideration has occurred to me. From the programming point of view, what seems *not* to be going on is that I am operating with a huge number of IF THEN statements followed by an ELSE statement (infer *p*; do *A*), where each IF clause contains some adverse possibility which the corresponding THEN clause says how to deal with, when detected. On that model, if no condition specified by an IF clause is met, the ELSE clause goes into effect and I infer *p*. But subjectively what seems to happen is that I operate with a statement that says to infer *p* UNLESS . . . , where the blank is filled in by all the IF THEN statements. It may be logically equivalent to the complex IF THEN ELSE statement, but it seems to function differently, in that I don't *seem* to check for the fulfillment of the UNLESS clause. The situation is analogous to what happens when I carry out a categorical intention. As noted earlier, such an intention is not an intention to do something no matter what, but it is not a qualified intention either. That is, the conditions under which I would not carry it out are not specified in its content. So normally I don't check to see that no such condition materializes, but I do expect to notice if one of them does, in which case I would revise my plans.

I don't mean to suggest that there is nothing going on in me to check for contrary considerations. Maybe, as in certain AI programs, I have demons to do the job. That idea, introduced in AI by Charniak (1974), goes back to Socrates: "A sort of voice [*daimon*] sometimes comes to me. It never tells me what to do but only dissuades me from doing what I am proposing to do" (*Apology* 31D). Still, there is the question of how, in real-life situations, the right demon goes to work and how, moreover, I get access to its findings. I don't know how these demons are organized or how particular ones take control, but some explanation is needed of how cognitive processes that normally go quickly and smoothly are sometimes interrupted (see Pylyshyn 1979), and appropriately complicated, by special considerations that come to mind. The problem is not merely, as is often said, that we can be surprised or startled by something whose absence we did not explicitly expect, but also that this generally occurs just when it should. A similar phenomenon occurs in the context of skilled activity. Things are going along smoothly and then something doesn't feel right, and we pay special attention to what we're doing.

I cannot review the recent work in AI that is relevant here, but it is more promising than you might suppose, especially if you have been influenced by latterday phenomenologists like Dreyfus (1981) and Searle (1983). These critics

of AI claim that what they call the "background" (what people in AI recognize as generating the frame problem) undermines the representational/computational theory of mind. Even Jerry Fodor is pessimistic about the prospects of simulating global cognitive processes (1983, 129), partly because he suspects that nonpropositional systems of representation, such as Minsky's (1974) frames, are not genuine but merely notational alternatives to propositional systems (1983, 116). However, some of the recent work alluded to above (Lenat 1982, 1983 and Aikins 1983) imposes enough structure on its systems, including modifiable heuristics, as to suggest that these are genuine alternatives. Besides, I know of no good reason to suppose that mental representations *are* propositional, at least if that means that they are couched in some form of natural language like English or in some sort of formal language like the predicate calculus. Minsky's (1974) frames (not so-called because of the frame problem) or something like them seem much more plausible. Individual frames contain rich sets of presumptions, stereotypes, or default assignments, call them what you will. Moreover, frame systems are organized in such a way that inferences are selectively triggered and available information is selectively utilized.³³ If anything like the frame model applies to real people, it could make sense of the fact that in many situations we have certain expectations which are noticed only in their frustration. That is, we don't test our environment to see if our expectations are met; rather, we are startled when they are not. For example, when we enter a room, we expect the lights to go on when we flip the switch and the floor to support us as we cross it. Perhaps these expectations are due to default values of the frame (here the entering-a-room frame) being applied to the current situation. If we notice something inconsistent with one of those values, we are startled and a procedure is activated to explain why. We are startled because there is a clash between some expected value and the detected value of the same parameter. The checking procedure that is thereby activated gives rise to further expectations, given by the default values in the frame that represents the information about that procedure. These might be about the likely results of certain tests, whose actual results may activate still further procedures as represented by further frames.

I don't think, by the way, that the frame approach really threatens the ordinary notion of belief and other concepts of so-called folk psychology. Folk psychology contains no theory of mental representation. If something like the frame model is correct, then what we call beliefs are not represented one-by-one, contrary to what the language of thought hypothesis suggests. As a result, the frame model blurs the distinction between core beliefs and implicit beliefs, since it leaves unclear what it is for a belief to be explicitly represented. Having a certain belief may consist in a certain slot in a certain frame being filled in a certain way. However, it could just as well consist in the fact that the person's frame system works in a certain manner. For example, if, when a certain expectation is not met, he runs one particular test before running another, where there is no difference in cost between the two, that he runs the first might constitute his belief that what it tests for is more probable than what the other tests for.

In both thought and action we engage in default reasoning. This involves making implicit assumptions (they don't come to mind) and following rules like

the take-for-granted (TFG) and the not-worth-considering (NWC) rules. By following them we can make judgments and decisions directly or at least bypass some intermediate steps. We jump to conclusions (final or intermediate) when it does not occur to us not to. We rely on our ability to know when to look before we leap, without having to judge that we don't need to look (making that judgment would itself be taking just the sort of intermediate step that was to be bypassed). In bypassing intermediate steps we default to the conclusions we draw. Workers in AI, being forced to deal in detail with the informational demands of problem solving and planning, recognize the pervasive role of default reasoning and have tried various ways of modeling it.

Rules like TFG and NWC contain conditions requiring the nonoccurrence of certain thoughts. Following such rules is reasonable to the extent that the specified thoughts do occur when the default reasoning should be interrupted. To follow such rules is to exercise intelligent cognitive habits and skills (not that we can't make mistakes by force of habit). These habits and skills enable us to bring antecedent beliefs and intentions as well as newly detected information to bear on our ongoing processes of reasoning. The take-for-granted (TFGP) and the not-worth-considering (NWCP) principles, applicable to factual and practical reasoning respectively, illustrate the conception of rationality operative here, one that reconciles efficiency with reliability. In proposing these principles, I do not pretend to have offered an account of when it is reasonable for a given thought to occur or not to occur, but such an account is needed for a theory of human rationality. After all, real people, unlike philosophers, have to make judgments and decisions all the time, not just when they're good and ready.

San Francisco State University
San Francisco, California

NOTES

¹I leave open whether the common view of reasoning requires that "valid" means "deductively valid" (i.e. "truth-preserving"). It surely doesn't mean "formally valid," although some who hold a logistic view of reasoning might insist that reasoning ideally should exemplify a formally valid argument. But this would be to make the controversial assumption that thought contents can be fully formalized.

²This seems to be why Gilbert Harman rejects the logistic conception of reasoning. He argues that logic has far less to do with good reasoning than is commonly supposed, in that "so-called deductive rules of inference are not plausibly construed as rules of deductive acceptance." (1973, 157) In the case of *modus ponens*, for example, logic does not tell you, if you believe both *p* and if *p* then *q*, to conclude *q*. That conclusion is entailed by the premises all right, but perhaps the thing to do is reject one of those premises rather than accept *q*, which you may have independent reason to reject. Similarly, although logic says that anything follows from a contradiction, it does not tell you to draw conclusions galore, should you ever discover an inconsistency in your beliefs. Harman is certainly right about so-called rules of inference, but he has not shown that logic has little to do with good reasoning. Logic as we know it does not tell you which valid conclusions to draw and when to give up premises instead, but for all we know, perhaps there is a logic of doxastic conflict resolution.

³Later I found that Stuart Hampshire had earlier observed, though at less tender an age, that we are always doing something and "are therefore always actively following what is happening now as leading into what is to happen next" (1960, 119). And Allen Newell and Herbert Simon, at the end of their Turing Award Lecture (1975), cryptically remark, "the critical question is always: What to do next?"

⁴Relevant here is Herbert Simon's observation that people generally don't optimize but merely "satisfice" (1957, xxv), but he was talking mainly about motivation. People satisfice not merely because they are satisfied with less than the best but because optimizing is generally unfeasible. One reason for that is our cognitive limitations (in this regard see Christopher Cherniak (1979) on "feasible inferences" and Allan Gibbard (to appear) on "attainable rationality"), but there is also the ongoing problem of what to do next: new judgments and decisions keep having to be made. As we will see later on, this problem is endemic to logistic systems in AI.

⁵Of course it is not the business of philosophy to investigate the nature of these processes, but there is still something it can do. It can identify human competencies and argue that if they are not magic powers but psychologically explicable abilities, then certain subcompetencies might very well be necessary for having them. But just how the latter are realized is another question, one for cognitive science. Ultimately, we may be able to analyze abilities into basic subabilities, ones that are constituted by processes best described in nonpsychological terms (see Cummins 1983).

⁶I defend this claim in Bach (to appear), and use it to argue against the internalist conception of justified belief. According to internalism, if believing *p* is justified for *S*, then *S* must be aware, or at least be immediately capable of being aware, of what makes it justified and why. The idea seems to be that good reasoning requires that the person justify, or at least be able to justify, his reasoning every step of the way.

⁷This is analogous to William James's *ideo-motor* theory of action, according to which an "anticipatory idea" leads to bodily movement if it is not immediately inhibited by some contrary idea.

⁸Philosophers usually call it "theoretical," but that strikes me as too grandiose a label.

⁹Conversely, beliefs and intentions usually result from judgments and decisions—but not always. For as Stuart Hampshire observes, "Intentions, like beliefs, are not always and necessarily the outcome of a process of thought or of a datable act of decision. They may, like beliefs, effortlessly form themselves in my mind without conscious and controlled deliberation" (1960, 101).

¹⁰Curiously, we speak both of a step in an argument "following from" (not merely following) what precedes and of a person "following" (understanding and accepting) a step of an argument. I am suggesting that at the most fine-grained level of an argument, judging that something follows just is following it. This is, in effect, a generalization of Lewis Carroll's (1895) observation about *modus ponens*: a legitimate inference of *q* from *p* and if *p* then *q* does not require the further premise that if *p* and if *p* then *q*, then *q*.

¹¹I have in mind the ingenious but extraordinarily complex solutions proposed by Marshall Swain (1981), Peter Klein (1981), and Robert Shope (1983). Even if one of their analyses of knowledge is correct to the extent that the *analysans* is extensionally equivalent to the concept of knowledge, its sheer complexity makes it highly implausible as an explication of the concept.

¹²Harman (1973). In this section page references are to that book unless otherwise indicated.

¹³Harman (1973, 115; 1980, 151) rejects psychologism that relies on intuitions about justified belief, on the grounds that these intuitions not sufficiently "secure." I do not. In Bach (to appear) I suggest that ostensibly conflicting intuitions about justification are really intuitions about two different things: what it is for a belief to be justified and what it is for a person to be justified in believing. In any case, psychologism would be a marginally fruitful strategy if we could use our intuitions only about "the Gettier effect to decide when reasoning has occurred and what reasoning there has been" (47).

¹⁴Harman often seems too casual about such considerations. He has no qualms about saying things like, "If [inference in perception] would have had to be instantaneous, then inference takes no time" (21).

¹⁵Williams thinks "that principle P is more plausibly applied to such cases if detached from the notion of 'real reasons'" (1978, 259) and "if the notion of 'reasons for belief' were explained in terms of what a person would (or perhaps ought to) say if questions of justification were pressed" (263). Thus he rejects Harman's view (1973, 24–32) that reasons for a belief must explain the belief.

¹⁶The difference here is *not* the difference between conscious and unconscious thought or anything of the sort: an explicitly represented step need not be conscious—indeed, the whole reasoning need not be—and an implicit step is not represented at all in the reasoning, consciously or unconsciously.

¹⁷This is not meant as an analysis of the concept of implicit assumption. If it were, it would be guilty of what Robert Shope (1978) has identified as the "conditional fallacy" in contemporary philosophy.

¹⁸The fact that a step in an inference pattern is bypassed in the process of reasoning that instantiates that pattern does not mean that there is nothing psychologically real corresponding to that step. There

is no explicit element of the reasoning process corresponding to that step, but still the step is being made implicitly, by virtue of the underlying assumption that supports the abbreviated version of the inference.

¹⁹Examples involving action and communication may be found, respectively, in Bach (1978) and Bach and Harnish (1979, ch. 10). Memory provides a host of other examples, as in the recall of names, statistics, spellings, and the diversity of facts (biographical, historical, geographic, etc.) popularly known as "trivia." Harman suggests that here our judgments are "based on reasoning concerning the best explanation of present memories" (1973, 189). In my view this reasoning is largely implicit. For example, when I see someone for the first time in years and immediately recall his name, normally I don't think to myself, "I am good at remembering names. Therefore, the first name that seems right to me is likely to be his name." Rather, I just start thinking of the person as having that name. If I did have doubts about it, I might reflect on my ability to remember names and seek an explanation of why I might not know or remember the name of this person. I might think that he is not who I thought he was. But I do not affirm my reliability to myself in the normal case where I don't have doubts.

²⁰Perhaps Harman would accept this possibility, as suggested by his discussion of inference in perception, where he draws a distinction between "automatic inferences and inferences that require some attention" (183). Unfortunately, he does not explain what requiring attention involves. Does the perceiver have to *infer* that attention is needed and, if so, how does this occur to him? Or does he somehow just know when to pay attention, as TFG suggests?

²¹However, in a benign, naive realist world, where things appeared only as they are, I wouldn't need the ability to detect circumstantial abnormalities but also I wouldn't need to follow TFG. In a world like that I could acquire plenty of justified beliefs, plenty of knowledge, simply by following AIR. Here I am assuming a reliabilist conception of justified belief, on which justifiedness is partly an empirical matter (Bach to appear).

²²I am not assuming here any particular conception of what it is to follow a rule, but I am assuming that there is a legitimate sense in which a rule can be followed without being explicitly represented or applied. Notice that even where there is explicit rule-following, there must also be, at a more fine-grained level, implicit rule-following. As Robert Cummins (1983, 44–51) points out, not all rule-governed (or programmed) behavior can be like following a recipe. For example, in following a cooking recipe you probably aren't also following a reading recipe, and even if you did, surely you would not also be following a lower-level recipe for identifying letters.

²³This is the causal not the reasons "because." At this level we do not do something for the *reason* that it occurs to us to do so and it does not next occur to us not to. We don't even notice that.

²⁴However, being unintended need not make an action unintentional (Harman 1976, 433–4).

²⁵This is generally true but not in self-deception (Bach 1981), where there is motivated resistance to the activation of beliefs that would otherwise come to mind spontaneously.

²⁶Sometimes life gets too complicated for that. Busy executives make plenty of decisions but have their administrative assistants do what for the rest of us is the work of intentions: (1) recording decisions of what to do when, (2) enabling one to take these decisions into account in subsequent deliberations, (3) leading one to devise and take necessary steps prior to the time of the action, and (4) prompting one, when the time comes, to do what was decided on without first reconsidering it.

²⁷The notion of subjective probability may have its uses, as in decision theory, but if taken as degree of belief it is psychologically unrealistic, owing to the problem of combinatorial explosions (Harman 1980, 154–5).

²⁸Things are even more complex with desire. Philosophers seem to think of strength of desire as one-dimensional, but I think desire comes in at least six different dimensions of "strength": (1) motivational force, (2) importance conferred on the object of desire, (3) felt intensity, (4) frequency and persistency of coming to and staying in mind, (5) insatiability, and (6) resistance to second-order desires to get rid of the desire.

²⁹Among categorical intentions, Harman distinguishes positive from negative intentions (1976, 440). This distinction is not simply based on the difference between positive and negative actions (refrainings). Rather, the idea is that a positive intention plays a causal role in doing what is intended, whereas a negative one does not, either because it is the intention to refrain from doing something or because it is the intention to do what, in the normal course of events, one would do anyway.

³⁰The frame problem was introduced to AI by John McCarthy and Patrick Hayes (1969). Daniel Dennett (to appear) sees its relevance to philosophy and discusses it in some detail, but does not explicitly distinguish it from the various other problems mentioned below.

³¹Worth noting here is the *updating* problem, sometimes confused with the frame problem. It is illustrated by the fact that we have countless beliefs of the form 'that *p*-now' about enduring states of affairs (as opposed to beliefs of the same form about ongoing events, such as that the bell is ringing now). For example, I believe that Halley's Comet cannot now be seen with the naked eye, and I will continue to believe that for some time to come, although the reference of "now" will keep changing. What does this continuous updating involve? Harman (1973, 189ff) suggests that such tensed beliefs are continuously re-inferred until abandoned. That's a whole lot of re-infering, since we have countless such beliefs (this is why the seemingly trivial updating problem is anything but that in AI). Harman's suggestion may seem implausible psychologically but not if implicit inference is allowed for. Indeed, perhaps present-tensed beliefs about persistent states of affairs are updated by a process of implicit inference governed by the following *updating rule*:

(UPD) If *p*-now, then continue to infer that *p*-now, as long as no reason to the contrary occurs to me.

Applying UPD is not a matter of continuously making, at each successive moment, an explicit inference that *p*-now for each '*p*'. Rather, the time reference of each of my beliefs of the form that *p*-now automatically changes from one moment to the next, without there being any change in my belief *state* (Perry 1980). Thus UPD is a rule of implicit inference. And, since UPD adverts to reasons to the contrary, my use of UPD is warranted to the extent that I am sensitive to information calling for substantive reconsideration of the relevant beliefs.

³²Nonmonotonic logics are so-called because adding premises to an argument can diminish what can be proved from them. An entire issue of *Artificial Intelligence* was devoted to these logics, including versions by Drew McDermott and Jon Doyle (1980), John McCarthy (1980), and Ray Reiter (1980). Doyle (1983) has since reviewed the entire literature on the subject.

³³For example, there can be frames embedded in frames, and there are "frame-keepers" to pass control from one frame to another.

REFERENCES

- Aikens, J. S. (1983). "Prototypical knowledge for expert systems." *Artificial Intelligence* 20:163–210.
- Bach, K. (1978). "A representational theory of action" *Philosophical Studies* 34:361–379.
- Bach, K. (1981). "An analysis of self-deception." *Philosophy and Phenomenological Research* 41:351–370.
- Bach, K. (to appear). "A Rationale for Reliabilism." *The Monist*.
- Bach, K. and R. Harnish (1979). *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.
- Carroll, L. (1895). "What the tortoise said to Achilles." *Mind* 4:278–280.
- Charniak, E. (1974). *Toward a Model of Children's Story Comprehension*. Unpublished Ph.D. thesis, MIT.
- Cherniak, C. (1979). "Feasible inferences" *Philosophy of Science* 45.
- Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, Mass.: Bradford Books/MIT Press.
- Dennett, D. (to appear). "The frame problem of AI." In C. Hookway, ed. *Minds, Machines and Evolution*. Cambridge, Eng.: Cambridge University Press.
- Doyle, J. (1983). *Some Theories of Reasoned Assumptions*. Pittsburgh: Department of Computer Science, Carnegie-Mellon University.
- Dreyfus, H. (1981). "From micro-worlds to knowledge representation." In Haugeland (1981).
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, Mass.: Bradford Books/MIT Press.
- Gettier, E. (1963). "Is justified true belief knowledge?" *Analysis* 23:121–123.
- Gibbard, A. (to appear). "Attainable rationality in action: an analysis."
- Goldman, A. (1976). "Discrimination and perceptual knowledge." *Journal of Philosophy* 73:771–791.
- Hampshire, S. (1960). *Thought and Action*. New York: The Viking Press.
- Harman, G. (1973). *Thought*. Princeton, NJ: Princeton University Press.
- Harman, G. (1976). "Practical reasoning." *Review of Metaphysics* 29:431–473.

- Harman, G. (1978). "Using intuitions about knowledge to study reasoning." *Journal of Philosophy* 75:433-438.
- Harman, G. (1980). "Reasoning and explanatory coherence." *American Philosophical Quarterly* 17:151-157.
- Haugeland, J. (1981), ed. *Mind Design*. Montgometry, Vermont: Bradford Books.
- Klein, P. (1981). *Certainty: A Refutation of Scepticism*. Minneapolis: University of Minnesota Press.
- Lenat, D. (1982, 1983) "The nature of heuristics." *Artificial Intelligence* 19:189-249, 20:31-98.
- McCarthy, J. (1980). "Circumscription—a form of non-monotonic reasoning." *Artificial Intelligence* 13:27-39.
- McCarthy, J. and P. Hayes (1969). "Some philosophical problems from the standpoint of artificial intelligence." In B. Meltzer and D. Michie, eds. *Machine Intelligence 4* New York: American Elsevier.
- McDermott, D. and J. Doyle (1980). "Non-monotonic logic." *Artificial Intelligence* 13:41-72.
- Minsky, M. (1974). "A framework for representing knowledge." Reprinted in Haugeland (1981).
- Newell A. and H. Simon (1975). "Computer science as empirical inquiry: symbols and search." Reprinted in Haugeland (1981).
- Perry, J. (1980). "A problem about continued belief." *Pacific Philosophical Quarterly* 61:317-332.
- Pylyshyn, Z. (1979). "Complexity and the study of artificial and human intelligence." Reprinted in Haugeland (1981).
- Raphael, B. (1976). *The Thinking Computer*. San Francisco: W. H. Freeman.
- Reiter, R. (1980). "A logic for default reasoning." *Artificial Intelligence* 13:81-132.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Shope, R. (1978). "The conditional fallacy in contemporary philosophy." *Journal of Philosophy* 75:397-413.
- Shope, R. (1983). *The Analysis of Knowing*. Princeton, N.J.: Princeton University Press.
- Simon, H. (1957). *Administrative Behavior*, 2nd edition. New York: The Free Press.
- Swain, M. (1981). *Reasons and Knowledge*. Ithaca, N.Y.: Cornell University Press.
- Williams, M. (1978). "Inference, justification, and the analysis of knowledge," *Journal of Philosophy* 75:249-263.