analyst I may learn things about my beliefs by developing a theoretical model of myself. Yet again, cultural factors may strongly influence my conscious beliefs about myself. I may tell myself I am not biased against women because my culture informs me I ought not to be, while the arguments I make at, for example, a selection committee suggest to observers the presence of a strong bias in my underlying psychological state.

My second methodological comment concerns terminology. I believe that statements such as that children "think" such and such, or "believe" such and such, should be made only with considerable caution. What do we mean, for example, when we say, as Gopnik does, "3-year-olds believe that cognitive states come in only two varieties"? The child can classify its states into two varieties, certainly, but "belief" that there are only two suggests a different order of cognitive skill. In the interests of clear thinking, it seems desirable in such circumstances to avoid using words like "belief."

Common sense and adult theory of communication

Boaz Keysar

Psychology Department, University of Chicago, Chicago, IL 60614
Electronic mail: *boaz@speech.uchicago.edu*

[Gop] Gopnik presents an important and provocative thesis – that intentionality is a theoretical construct developing early in childhood. Even though common sense tells us that we have direct access to our psychological states, and even though we experience this access as unmediated, it is not different from the way we gain our knowledge about the psychological states of others. Both are mediated by inferential processes. Our common sense, then, is a theoretical construct that leads us to make such a distinction. Whereas the target article focuses on supporting this claim with evidence from the developmental literature, I will instead concentrate on adults.

What is the adult's theory of intentionality? In contrast to the elaborate body of studies that directly evaluate the child's theory of mind (e.g., Astington et al. 1988; Gopnik 1990; Wellman 1990), the adult theory of mind has not been directly evaluated in Gopnik's paper. Some studies that she mentions can address this question (e.g., Nisbett & Wilson 1977), but there is no systematic attempt to describe an adult "theory of mind" as such. In this regard some philosophical theories are basically elaborate accounts of adult common sense (e.g., Searle 1980). The resulting theoretical notion of adult common sense is thus overly based on intuitions. In the target article, then, there is an assumption that we as researchers have direct access to adult common sense.

Given this assumption, common sense may still reflect one of the following: It could be that our belief in direct access to our psychological state of intentionality is a theoretical construct, or it could be the case that first-person experience of intentionality is indeed accessed directly. To distinguish between these possibilities Gopnik suggests two lines of argumentation. First, a continuity argument: Because we are the same entities, only older, we probably have the same theory of mind that we developed as children. Second, an argument by analogy: We are fooled into believing that our sense of intentionality is direct and not theoretically mediated in the same way that experts are led to believe that they do not use inferences but simply "see" solutions. Both arguments are suggestive; neither is conclusive. It is possible that after the child develops a theory of mind, as described in the literature, further maturation may involve first-person access to intentional states, access that is not theoretically mediated. One can imagine how such a change may even be prompted by the implicit theory of mind itself. These two alternatives may also be empirically indistinguishable. It could be that intentionality is indeed a theoretical construct but so deeply ingrained that its products perfectly mimic those of nontheoretical, direct access to psychological states. This is not necessarily a problem for Gopnik's account. Instead, it suggests that it would be interesting to evaluate the theoretical status of adult common sense systematically, and that such an investigation could provide a direct test for the kind of arguments the target article puts forth.

Gopnik suggests that a possible source of evidence for the adult theory of mind may come from situations that lead to errors, as in some cases of experts' failure. Communicative behavior may be a prime example of a relatively complex activity that promises to be revealing about the nature of the adult theory of mind. According to the target article, one aspect of a theoretically driven common sense is the conjunction of two elements: (1) Self-experience is interpretative and not direct, and (2) we believe that self-experience is direct and not interpretative. Because the comprehension of utterances is interpretative by its very nature, language use may be a reasonable place to look for relevant evidence.

Some preliminary studies in language use may be consistent with Gopnik's claim about the adult's theory of mind. For example, it seems that speakers may not be aware of the interpretative nature of their own utterances. Elizabeth Hinkelman (personal communication) found that speakers are quite poor at reconstructing their own speech acts when they observe a videotape of a spontaneous interaction. Even when they are provided with the complete context, they are not always sure whether their utterance was a request, a suggestion, or a hint. Similarly, Anne Henly has collected data in my lab suggesting that speakers are not much better than chance in their attempts to identify their own intentions when they produced syntactically ambiguous sentences. For example, speakers who described a picture as "the man is chasing the woman on a bicycle" were not always able to tell whether they meant that the man or the woman was on the bicycle when presented with their own utterance the following day. Such examples suggest that when speakers are producing their own utterances they may not be fully aware of their interpretative nature. Similarly, when people are asked to evaluate another person's interpretation of utterances, they seem to rely on their own understanding as if it were not interpretative: When people understand an ambiguous utterance as sarcastic (e.g., "Thank you for the helpful advice") they believe that others would perceive the same intention even when they know that the others lack information that is crucial for the perception of sarcastic intent (e.g., information that the advice was not very helpful). It may be that they perceive their own interpretation as "direct," consequently underestimating the inferences that led to the perceived intention (Keysar 1991). In other words, they take their understanding to be noninterpretative.

Because adult theory of communication and language use is a subset of the general theory of mind, it provides a reasonable place to start. Whether or not the results will support a theory of mind in line with the arguments of the target article, Gopnik's target article does challenge us to investigate adults' theory of mind systematically and directly.

Self-attributions help constitute mental types

Bernard W. Kobes

Department of Philosophy, Arizona State University, Tempe, AZ 85287-2004
Electronic mail: *atbxk@asuacad.bitnet*

[Gol, Gop] People have both (a) "object-level" sensations and beliefs and desires about the world and (b) ordinary practices of describing and explaining themselves and others mentalistically (folk psychology in Goldman's broad sense). What is the relation

between the two? I shall argue that folk psychology, especially insofar as it includes self-attribution, is a special practice of description and explanation that helps constitute its own subject matter.

Goldman is a bit unimaginative in the way he saddles analytic functionalism (AF) with psychological baggage. He criticizes AF on the grounds that it ought to try to be psychologically realistic, in the manner of representational functionalism (RF). Now AF tries to distill out of ordinary (i.e., nontechnical, not philosophically self-conscious) usage and beliefs a network of causal interrelations that is definitive of each mental state's type-identity. Articulating these is an enterprise distinct from the psychology of folk psychology. It is conceivable, for example, that ordinary people make self-attributions directly on the basis of nondefinitive, heuristic criteria that are reliable often enough to be worthwhile but go systematically wrong under adverse conditions. AF may accommodate this in any of three ways. First, the more articulated functional roles that (according to AF) define mental-state types may be stored in long-term semantic memory but not be ordinarily used except as a fall-back when there is reason to double-check. Second, definitive functional roles may not be stored in the minds of many ordinary speakers at all, but only in the minds of some speakers to whom the many defer. AF may postulate a semantic division of labor. Third, definitive functional roles may be stored in a highly implicit and generalized form (I shall return to this point). So folk psychology need not explicitly represent causal relations definitive of mental types and Goldman's assimilation of AF to RF seems, even from a broadly naturalistic perspective, something of a straw man.

Shoemaker (1975) argues that it is part of the very functional specification of pain (say), that it causes a belief that one is in pain. Thus, according to Shoemaker's view, in effect, a self-attribution of pain helps constitute the state it is about as one of pain. Does Goldman's schematic RF allow for this? Suitably construed, it does. On RF a belief that one is in pain occurs when a match occurs between a category representation (CR) for the mental state and a suitable instance representation (IR). Goldman infers from this that one could only come to believe one was in pain if one already believed one was in pain. But why should we acquiesce in the assumption of temporal priority? If coming-to-believe that one is in pain helps constitute the state one is in as one of pain, then we would expect the mental state and the self-attribution to be concurrent. This is not inconsistent with the matching model if the process of matching is allowed to include itself as part of the pattern being matched.

Compare the following piece of reasoning, which I shall call (R). Premise: (R) is a valid argument. Therefore, at least one valid argument exists. Evidently (R) is itself a sound, albeit self-referential, argument. Analogously, a mental match may take place which includes itself as a component of one of the items matched. Moreover, it is plausible that explicit noncircular models for the matching procedure could be constructed, especially if, as Goldman allows, a partial match may trigger an initial type identification of the mental state, which then receives a measure of "bootstrapping" confirmation from the type identification itself.

In my view we should be idealistic about the mind, not in the trivial sense that what is mental is mentally constituted, but in the more substantial sense that object-level mental states and events, the subject matter of folk psychology, are partly constituted by our self-attributions and hence indirectly by our ordinary public practices of mentalistic description and explanation. The view is not that whether I am in mental state *M* or not is indeterminate or nonfactual, nor that the question is somehow up to me to decide (the relevant self-attribution may be involuntary and automatic). It is rather that psychologists cannot assume that "object-level" mental-state tokens belong to determinate types prior to and independently of relevant self-attributions.

Goldman suggests that the Schachter and Singer (1962) data might best be construed as showing that cognitive influences help determine which emotion is actually *felt*, rather than merely the process of labeling or classifying the felt emotion. If so, then a self-attribution of anger (say) may help constitute the relevant emotion-instance as one of anger. This example also illustrates the impossibility in certain cases of distinguishing between phenomenological awareness and awareness of functional role. Goldman himself allows great latitude and variety in the objects of conscious awareness: Attitude types such as doubt and disappointment, even propositional-attitude contents, may serve. Might we not, then, be consciously but implicitly aware of functional roles? In being aware of a propositional-attitude content p we are implicitly aware of how coming to believe p would change our current beliefs and desires. We are also aware of the strengths of propositional attitudes such as belief and desire, and this too constitutes an implicit awareness of an aspect of functional role. Goldman endorses recent psychological work according to which pain has three microfeatural dimensions: character, intensity, and aversiveness. But aversiveness, how much the subject minds the pain, is surely a functional aspect of the experience, concerning the strength of its causal connection to actions seeking to diminish the sensation. So, given Goldman's own views, it is plausible that we can be consciously but implicitly aware of functional roles. Some phenomenological qualities may not be monadic but relational, in the manner of functional states.

This raises the question whether Goldman's phenomenological model is properly characterized as an alternative to functionalism. If the phenomenological quality of a mental state or event may count as conscious awareness of functional role, then that functional role may be matched against the functional role represented in the CR. Thus, when I wake up with a headache, I am quickly aware of being in a state that will tend to cause me to take steps to get rid of it – aware, that is, of its aversiveness. I have reliable and fast information about its likely behavioral effects under various possible circumstances. Of course, I do not represent each of the infinitely many possible counterfactual situations discretely and explicitly, but I do represent them in a generalized, unified form, and this may be sufficient to trigger a quick match with the CR for headache. So this approach makes headway against Goldman's arguments from the ignorance of causes and effects and of subjunctive properties.

We can also defuse the threat of combinatorial explosion, for in order to make the match between CR and IR the subject need not explicitly and discretely represent each of the other mental states to which the IR is causally linked. These countless states need be represented only once in long-term semantic memory as part of a single theory of mind, either explicitly or implicitly via some generative structure for propositional contents. The IR may be an experiential representation of functional role. Moreover, although the functional role of the instance includes a wealth of causal connections, many of these may be causally mediated by the very match between IR and CR and the concomitant self-attribution.

Consciousness is a sufficiently peculiar phenomenon that, antecedently, we should not be surprised if it turns out to be reflexive in the way I have sketched. The model is surely less mysterious than the direct detection of (nonphysical?) phenomenological properties to which Goldman apparently subscribes.

I want to close with some remarks about Gopnik's theory of first-person knowledge. I find much of what Gopnik says under this head plausible and I will assume that our theory of mind may have a bearing on the particular attributions we make, even self-attributions. But it is difficult to accept that the apparent authority of strictly present-tense self-attributions is just an illusion due to expertise. So I want to suggest an answer on Gopnik's behalf to Goldman's challenge on this score: "If faulty theoretical inference is rampant in children's self-attribution of past states, why do they not make equally faulty inferences about their current states?" (sect. 10, para. 8).

For attributions to others we have only external evidence, but for our recently past selves we have a higher grade of evidence, namely, fresh memories of inner states Three-year old children too have, presumably, good memories of the recent past. So it is *not* plausible that memory decay accounts for children's systematically erroneous beliefs about their immediately past beliefs. Rather, as Gopnik argues, the child's theory of mind prevents the child from giving this memory evidence its proper weight, or perhaps from interpreting it correctly. Yet it is remarkable that nobody of any age makes similar errors about strictly concurrent mental states. Why is this? High-grade evidence cannot be the whole story, for it is almost as good for recent-past self-attributions, even at age 3, · et strikingly insufficient in that case.

I have argued that self-attribution helps constitute the type-identity of the relevant object-level mental state – but plainly this holds only when the self-attribution is present-tensed. If a mental state is in the (even very recent) past, then its type is already fixed, regardless of subsequent self-attributions. But present-tense self-attributions help constitute the type-identity of the relevant object-level states in such a way as to tend to make them self-verifying. This is so even if all self-attributions are as theory-laden as Gopnik thinks. So if a 3-year old (perhaps in consequence of theory-laden self-cognition) sincerely says or thinks that he thinks that there are pencils in the box, or that he wants chocolate mousse, then those self-attributions help constitute his present mental state as one in which he really does think there are pencils in the box, or really does want chocolate mousse. So this model yields an attractive account of the contrast between present-tense self-attributions and recent-past self-attributions in 3-year-olds.

Even a theory-theory needs information processing: ToMM, an alternative theory-theory of the child's theory of mind

Alan M. Leslie, Tim P. German and Francesca G. Happé

MRC Cognitive Development Unit, University of London, London WC1H 0AH, England

Electronic mail: *a.leslie@ucl.ac.uk*

[Gol, Gop] Although we endorse the theory-theory view in general terms, we think the specific "consensus" version Gopnik advocates is wrong. We do not believe that the preschool child's success on false-belief (FB) tasks reflects the construction of a representational theory of mind (RTM), nor do we believe that the child's theory of mind undergoes a radical conceptual shift around the age of 4.

Gopnik endorses the consensus RTM view of preschool development that is seen at its most explicit in Perner (1991b). The key notion is that success on standard FB tasks at 4 years is the result of a conceptual shift to RTM. The vital question for any developmental theory-theory is where the theory and its concepts come from in the first place. To the limited extent that advocates of preschool RTM address this question, the following answer can be gleaned. Because FB is a misrepresentation of a situation in the world, it can be understood by the child only in terms of a theory of representation. The child constructs a theory of representation by (somehow) learning about artefacts like pictures (models, maps, etc.), which being both public and observable are easier to learn about than beliefs. Having thus developed a theory of representation, the child applies it to the mind in the form of a pictures-in-the-head theory of mental states. Therefore, understanding public representations should occur earlier than understanding FB.

The above story can be given sophistication: Although a photograph is a representation, it cannot be false in the way a belief can be false. A photograph is simply an accurate representation of a situation (e.g., the chocolate sitting in the cupboard). If the situation changes, the photograph is simply a still-accurate representation of the old situation, not a *mis*representation of the new situation. In the FB task, Maxi's belief starts, like the photograph, as an accurate representation of the situation. When the situation changes, however, unlike the photograph, Maxi's belief does become a misrepresentation of the new situation. This is because Maxi mistakenly *believes* his representation (of the previous situation) accurately represents the current situation. The photograph cannot perform this trick because the photograph cannot believe anything.

Notice that the difference between the two cases above is precisely related to the special nature of *believing* (and more generally, to the nature of propositional attitudes) rather than to the general problem of the nature of representations. Understanding representation then could only be a subcomponent of understanding belief. Unlike the photograph itself, Maxi could mistakenly believe that the photograph depicts a current situation. In this account, the problem of understanding representations, like photographs or pictures-in-the-head that go out-of-date, is included as a subcomponent in the problem of understanding beliefs that go out-of-date. False belief includes all the conceptual complexities of representational pictures plus some other complexities specific to belief. Again, reinforced by the idea that public representational artefacts will be easier to learn about, this predicts that out-of-date pictures will be understood earlier than FBs.

Unfortunately for this account, the evidence from preschool development clearly contradicts the prediction. When tested in the same way, FB is reliably easier and "understood" earlier than pictures (Leslie & Thaiss 1992; Zaitchik 1990). Either (a) one must find an analysis in terms of general processes of theory construction in which FB is *less* complex than out-of-date pictures or (b) one abandons the assumption of purely general processes and looks for an account of FB understanding in terms of specialized, domain-specific mechanisms. If one opts for the first of these, one cannot account for the performance of autistic children, which is near ceiling on out-of-date pictures but severely impaired on FB. This leaves the second opinion to which we return below.

Information processing and theory of mind. Gopnik has an idiosyncratic notion of what an information-processing theory should be. She stipulates (sect. 5) that an information-processing account of the shift in performance on FB tasks between 3 and 4 years of age must advert only to a single factor in explaining the differences. Despite Gopnik's worries about parsimony, this stipulation seems entirely arbitrary. A task analysis may well reveal that a number of specific and nonspecific mechanisms are involved across theory-of-mind tasks. The nonspecific mechanisms will also be involved in tasks outside theory of mind. Actually, Leslie and Thaiss (1992) do propose a nonspecific problem-solving mechanism that might neatly divide "easy" from "difficult" theory-of-mind tasks. However, this does not commit anyone to there having to be one, and only one, such component involved.

Gopnik discounts "information processing" largely because she conflates the rejection of the "shift-at-4" hypothesis with rejection of the theory-theory. But the former is only one version of the latter position; not all theory-theories need espouse the "shift-at-4" hypothesis. Indeed, is not attributing a single theory to both 3- and 4-year olds more parsimonious than two theories plus a shift?

Finally, Gopnik assumes that an information-processing account is on a par with a theory-theory, that is, it would compete with a theory-theory. But an information-processing account is necessary whatever theory-theory one adopts (and indeed for any simulation theory too). Information processing provides the framework for cognitive science and thus for particular theories of cognitive abilities. Does Gopnik really believe that her theory-theory does not assume the processing of information?