

Chapter 5

Whose Concepts Are They, Anyway? The Role of Philosophical Intuition in Empirical Psychology

Alison Gopnik and Eric Schwitzgebel

A Dialogue in the Sandbox

Over the past ten years, interest among developmental psychologists in how young children conceive of the mind has grown dramatically. A very generally, though not universally, accepted hypothesis is that there is a radical conceptual change in children's understanding of belief between about three and five years of age. Three-year-olds seem to believe that all beliefs are true, and if they know that the world is a particular way, they seem to think that no one else could believe otherwise. Four-year-olds, in contrast, seem to have a view more like the adult view. They know that beliefs can be false, can differ between people, and can change over time, even while the reality behind the beliefs remains constant. The prevailing view suggests that this development happens in part as a result of children's experience of the behavior of others. Behavior that seems inexplicable without the assumption of a false belief on the part of the agent, for example, can serve as counterevidence to the three-year-old's understanding of belief.

One day the first author of this chapter spent all morning in the daycare center talking to three-year-olds about belief and then all afternoon talking to a distinguished visiting philosopher about the same subject. That night she had a strange dream, a dream composed, as dreams often are, of an odd reshuffling of the day's events. In the dream, she happened to overhear the conversation of two three-year-olds in the daycare center sandbox. We will call them Phil and Psyche.

Psyche: You know, Phil, something's been bothering me. You know how beliefs are always true? Well, an odd thing happened the other day. My big brother saw my mom put a piece of chocolate in the cupboard and then left to play Nintendo, and while he was away my mom took the chocolate out of the cupboard and put it in the drawer. When my brother came back, he went straight to the cupboard and said loudly, several times, that he was sure the chocolate was in there. But of course, it was really in the drawer. So I have this idea: Could it be that he had a belief that was just like ordinary beliefs, except false?

- Phil: My dear Psyche, as I have so often pointed out to you before, your confusion is due to a category mistake. You are treating the truth of beliefs as if it were an empirical matter. Actually, it is simply a conceptual fact about beliefs that they are always true. Indeed, we might say that it is criterial for a belief to be a belief that it be true. Look, consult your intuitions, consult the intuitions of anyone else in the sandbox. All of us agree, immediately, intuitively, without inference or theory, that all beliefs are true. Ask yourself what a belief is. What else could it be but a true representation of events?
- Psyche: But couldn't we all be wrong? Couldn't there be an alternative way of conceiving of belief that none of us happen to subscribe to now?
- Phil: Another category mistake. When I say that beliefs are necessarily true, this isn't a mere contingent psychological fact about the concepts of all us three-year-olds. It's an eternal, platonic, *philosophical* fact about the nature of belief and truth.
- Psyche: Well, what about my brother?
- Phil: He is probably participating in an alternative form of life. I always thought he was kind of weird.
- Psyche: But you see, it isn't just him. It even seems to be me. Since the chocolate incident, wherever I look, I see evidence that beliefs may be false. Why just yesterday, a woman came into the daycare center with a candy box and I said "Candy!" and then she opened the box and there were pencils inside. I know intuitively that I must have thought there were pencils in the box all along, and of course that's what I told her when she asked me. But then why did I say "Candy!"? Am I turning into a madwoman?
- Phil: (gravely) I fear you may have a worse affliction. I fear you are turning into a cognitive psychologist. As I was saying just the other day, "It would be dangerous to deny from a philosophical armchair that cognitive psychology is an intellectually respectable discipline, provided, of course, it stays within proper bounds."¹ This is what happens when those bounds are breached.
- Psyche: But surely there must be some explanation?
- Phil: Philosophy does not provide explanations, only diagnoses. (Intones) Of that we cannot speak, thereof we must be silent. . . .

At this point the dreamer woke up suddenly, if not quite screaming, at least gravely shaken. She was temporarily comforted by the thought that in the sandbox kids like Phil get thwacked over the head with a bucket and the cognitive development of three-year-olds as a group proceeds unchecked. But in spite of appearances, we do not want to thwack our philosophical readers over the head with a bucket. Instead, what we would like to do is suggest some ways of making the dialogue between philosophers and psychologists more productive than it was in the dream. We would also like to point out some arenas where that kind of productive dialogue is already taking place. Philosophical work has a real contribution to make to the enterprise of understanding the nature of the mind, a contribution that is sometimes under appreciated by cognitive psychologists. At

the same time, we want to suggest that some empirical results in psychology, particularly developmental psychology, may have important implications for philosophical research, and we want to explain in some detail the rather common philosophical error of which we think Phil is guilty.

The core of Phil's argument is his appeal to the intuitions of the children in the sandbox. This chapter will explore both the benefits for empirical psychology of such philosophical appeals to intuition—the intuitions not of children but of educated adults—and the limitations of such appeals. We will examine several ways in which philosophical attention to intuition can contribute to empirical scientific psychology. We will then discuss Phil's error and describe some areas in contemporary philosophy where we think similar errors are being made. Finally, we will provide developmental evidence for the mutability and fallibility of our everyday intuitions about the mind, evidence that undermines arguments, like Phil's, that depend on taking such intuitions as a final authority for substantive claims about what the mind is like.

The Nature of Intuition

It will be helpful to begin by making clear what exactly we mean by the word 'intuition'. We will call any judgment an *intuitive judgment*, or more briefly an *intuition*, just in case that judgment is not made on the basis of some kind of explicit reasoning process that a person can consciously observe. Intuitions are judgments that grow, rather, out of an underground process, of whatever kind, that cannot be directly observed. So, for example, we make intuitive judgments about such things as the grammaticality of sentences, the morality of actions, the applicability of a certain term to a certain situation, Bob's likely reaction to an insult, the relative sizes of two distant objects, and so forth. In each of these cases, the judgments flow spontaneously from the situations that engender them, rather than from any process of explicit reasoning.

The content of these intuitive judgments may vary. In particular, we may have intuitions about aspects of the physical world, for example, we may intuit that objects cannot be in the same place at the same time, or that an object that is moving will end up in a different location than where it started, or that causes must precede effects. We may also have intuitions about the psychological world, we may intuit that pain always involves an element of displeasure, or that mental states like belief refer to the outside world.

Much philosophical work consists of articulating and elaborating these intuitive judgments. Sometimes philosophers elaborate on our ordinary intuitions about the physical world. John Mackie (1974), for example, has explored and criticized our intuitive view of causation, including such ideas as that a cause must be something without which the effect would not have occurred and that a cause cannot come after its effect. John Campbell (1995) elaborates on our ordinary conception of space and time. More often, philosophers elaborate on intuitions about the psychological world. Daniel Dennett (1978), for example, has explored some of the consequences and difficulties of the intuitive view that pain, by its nature, has an element of displeasure. John Searle (1983) presents an

elaborate intuitive taxonomy of our mental states. The views these philosophers describe are intuitive in the sense that ordinary people (or at least ordinary English speakers) frequently make intuitive judgments that agree with the views and would, on reflection, generally assent to those views, even if they could not support their assent by any explicit argument.

Intuitions, are, surprisingly, not always obvious. Philosophers may go to some length to ferret out our intuitions on a particular topic. Hilary Putnam (1975), for example, constructs an elaborate scenario involving Earth and a planet, "Twin Earth," identical to Earth in every respect except that on Twin Earth there is no water, only "twater," identical to water in every respect observable to the inhabitants of Earth and Twin Earth, but composed of chemicals other than H_2O . If Wayne on Earth and Dwayne on Twin Earth are molecule-for-molecule identical to each other (except that Dwayne is 70 percent twater), it is intuitive to claim that they, nonetheless, mean different things when they utter the word 'water'. This intuition is generally taken to support the surprising thesis that the meaning of words depends not wholly on what takes place in one's head but also on one's environment.

The Contribution of Philosophical Intuition

We suggest that there are several quite different ways in which philosophically-informed intuition can contribute to empirical psychology. It may be that other applications will occur to the reader—we do not mean to suggest that our three applications exhaust the possibilities.

Intuitions as Hypotheses

Perhaps the most obvious application of intuitions to the empirical sciences in general, and of philosophical intuitions to psychology, in particular, comes when we use our intuitions as a source of empirical hypotheses. We might think of intuitive judgments as particularly plausible initial hypotheses about the nature of the world. Thus, our intuitions that pain has an element of displeasure and that causes cannot come after effects, lead to the hypotheses that, in reality, pain has an element of displeasure, and that causes cannot come after effects.

Like all such hypotheses, our intuitions are accurate as a guide to the world to different degrees, depending on both the person who has those intuitions and the subject area to which those intuitions apply. Twentieth-century physics, for example, has notoriously shown the inaccuracy of our everyday intuitions as hypotheses about the nature of space and time, light, and the microscopic world. The well-cultivated intuitions of a theoretical physicist may be more accurate in this regard, to the extent her physical theories work their way underground into her intuitive judgments. In contrast, our commonsense intuitions about the behavior of mid-sized objects, as we interact with them in everyday life, are generally rather good, though there are notorious exceptions. For example, many adults intuit that objects that are thrown in a curve will continue moving on a curved trajectory until they complete a circle. As another example, consider that

many of us have essentially no intuitions about the quality of different chess moves—we can only evaluate moves by reasoning about their likely outcomes—while Gary Kasparov surely has quite accurate intuitions in this regard. Our intuitions can develop substantially as our expertise grows.

We are all experts, in a pragmatic sort of way, in everyday psychology, and for this reason our everyday psychological intuitions can be taken as plausible initial hypotheses about how the mind really works. Empirical psychology can then use these everyday intuitions as a starting point for its research. The work of philosophers of mind like John Searle who articulate the everyday intuitive view of the mind can therefore aid empirical psychology by making explicit what the intuitive starting point is, and by providing a set of hypotheses from which to begin. Similarly, philosophical work that clarifies our intuitive views of semantics or epistemology can be employed as a starting point for empirical work in semantics and in understanding human knowledge. In order for philosophical work of this sort to be helpful, it should be noticed, it is not necessary that philosophers have better intuitions about the mind, or semantics, or knowledge than the rest of us do. It is simply necessary that philosophers be skilled at spelling out the intuitions they (and we) do have.

The intuitions articulated by philosophers, then, in so far as they are treated as initial hypotheses about the way the world really is, can guide empirical research, especially in those domains in which human beings have particular expertise. Indeed, it is probably impossible to start a science from scratch with hypotheses and assumptions that are entirely based on observation or experiment. One must begin with intuition and correct it with experiment. As sciences mature, however, they typically revise, alter, and sometimes entirely reject, these initial hypotheses. In Galileo's day, it was permissible for physicists in defending one view or another to appeal to what we would now call our "folk physical" intuitions about what would happen in various circumstances. In contemporary physics, such appeals would be ruled out.

Similarly, as psychological science has matured, we have become more confident in leaving intuitions behind when they conflict with well-supported empirical findings, as for example in the cases of blindsight (Weiskrantz 1986) and attribution error (Nisbett and Ross 1980). Our initial hypotheses may be, indeed often are, moderated or defeated by later evidence. So although we may justifiably take reflective psychological intuition as a good preliminary guide, we no longer take it as the final authority about the mind.

Intuitions as Evidence

We can also use intuitions in a quite different way. Sometimes we treat intuitions as evidence about the nature of the mind of the person who generates those intuitions, not as hypotheses about the external world. In these cases we think of intuitions as data that need to be explained by a psychological theory. Grammatical theories, for instance, generally regard our linguistic intuitions about the grammaticality or ungrammaticality of sentences as a crucial part—even, perhaps, the whole—of their explanatory domain, just as celestial mechanics takes the movement of the stars and planets as a crucial part of its explanatory domain.

The intuition that the sentence “Jamie of gave” is ungrammatical is a piece of data for grammatical theories in roughly the same way that the observation that Venus was at such-and-such a location at such-and-such a time is a piece of data for celestial mechanics. Both the linguistic intuition and the astronomical observation can serve as evidence in support of or against particular linguistic and astronomical theories.

Yet the use of intuition in these cases should be differentiated from the use of intuition to provide plausible hypotheses. These intuitions are not first passes at a psychological theory, instead they are the very data that a psychological theory seeks to explain. Consequently, a theory that deviates from these intuitions must be able to provide some principled explanation of why the deviation should not be regarded as fatal to the theory; and a theory that deviates too much from the data it seeks to explain is doomed. When we consider them as evidence, intuitions have some of the same quality of indefeasibility that evidence always has. Parallel lines may, or may not, meet, and we may, or may not, have first-person access to our mental states, but it is indubitably true that we believe that parallel lines will not meet and that we have first-person access to our own mental states. A psychological theory has to explain why we have these intuitive beliefs even if, indeed especially if, the beliefs are quite false. The psychologist cannot simply reject the beliefs if she is treating them as evidence about the mind of the person who has them.

On the other hand, nothing stands in the way of the complete overthrow of intuitions if they are construed as psychological hypotheses. We might think of the contrast thus: Gary Kasparov’s intuitions about chess can be taken as *prima facie* support for one or another theory of the quality of chess moves, that is, they can be taken as hypotheses about chess, or they can be treated evidentially as a topic of study in themselves. In the former case, it may turn out that those intuitions must largely be repudiated, since the intuitions may not capture the facts about the quality of chess moves very well. Perhaps Kasparov’s hypotheses about chess moves are really not very good, say, if Kasparov has won by incredible luck, or only because everyone else is even worse at evaluating chess moves. But if we are trying to understand not chess, but Kasparov, the case is quite different. In that case, the question of the accuracy of the intuitions is only of secondary importance, since the piece of the world that the theory seeks to explain is not chess itself but Kasparov’s intuitions about chess.

Psychology frequently treats intuitive judgments as evidence in the construction of psychological theories. Perceptual psychology, for example, seeks to explain, among other things, our intuitive judgments about size, color, distance, shape, and so forth. Psychologists interested in science education have attempted to describe and explain our naive intuitions about physics. Developmental psychologists have attempted to characterize and explain children’s intuitions on a variety of subjects throughout childhood. Intuitive judgments are an important part of our mental lives; psychological theory, therefore, aims to explain the nature and origin of those intuitions.

Philosophical work that articulates our everyday intuitions on a particular topic can be useful for those parts of empirical psychology that take intuitions about that topic as their subject matter. Developmental psychology in particular

has greatly benefited from the philosophical articulation of our intuitions about a variety of subjects. In the literature on the development of the child's understanding of mind, the debt is often quite explicit. Developmental psychologists have mined intuitive characterizations of the mind by philosophers for material on which to test children: When do children understand that belief is "representational" in Fred Dretske's (1988) sense? When do they understand the relation between intention and action in roughly the way it has been articulated by philosophers of action? (Interestingly, it has been more typical for developmental psychologists to turn to philosophers for their characterizations of the adult conception of the mind than for them to turn to empirical research on the same topic.) Similarly, developmental psychologists studying the child's understanding of "natural kinds" have explicitly made reference to philosophical work on the topic (see e.g., Keil 1989; Gelman and Wellman 1991).

Theoretical Applications of Intuition

A third way in which intuitions articulated by philosophers can be applied to aid empirical psychology is in the elaboration of psychological theories and theoretical possibilities and in drawing out the consequences of these theories and possibilities. Used in this way, we will say that the intuitions are being employed *theoretically*.

Psychological theories, like all theories, are formulated in a language that relies on some mix of technical and ordinary terms. When the meanings and extensions of these terms are clear and univocal, the theories that employ them can be understood more precisely, and their empirical implications are more readily discerned. In the very best cases the concepts in the theories may be specified in very precise mathematical terms and the relations between them may be very clearly specified. Even in these cases, however, it may be difficult to tell if a mathematical concept is being applied appropriately. More frequently, however, the theoretical terms and concepts are not specified this precisely. When the meanings and extensions are left hazy, the theories become difficult to assess. Unless we really know what is meant by the term 'representation', for example, how can we fully evaluate a psychological theory that postulates the manipulation of representations in a certain functional subsystem of the brain?

Philosophical work that clarifies such central terms and concepts in psychology as 'representation', 'drive', and 'memory'—whether that work is done by philosophers attuned to psychology or whether it is done by psychologists with a philosophical knack—is plainly of great use to psychologists. When the terms and concepts in question are meant to be understood simply in their nontechnical, everyday sense, the articulation of intuitions about those terms should suffice to provide a profitably clear understanding of those terms for psychological purposes. Philosophers have long excelled at the sometimes difficult task of spelling out such intuitions. Intuitions develop quickly, also, among people using technical terms or appropriating everyday terms to technical uses, so that similar philosophical work can be done about the proper application of terms such as 'supervenience' or 'nominalist' in philosophy, or psychological terms such as those mentioned at the beginning of this paragraph. To arrive at a useful

understanding of such technical or quasi-technical terms requires a scientifically informed balance of intuitive considerations with historical, stipulative, and pragmatic elements. One might fantasize that the meanings of technical terms could simply be stipulated afresh each time they were employed, but of course in reality it is impossible fully to shed historical and intuitive associations. Regardless, any stipulative definition of a word must be provided in terms of other words, and it is difficult to imagine that the meanings of each of these words could be entirely stipulated, all the way down to the ground floor. Even in the case of highly mathematical and precise theories there is some point at which the theorist will have to rely on an intuitive judgment that, for example, a particular conclusion follows from certain premises, as in Lewis Carroll's paradox of Achilles and the Tortoise.

Philosophical work on the "theory theory" in developmental psychology can serve as a case in point. According to the theory theory of cognitive development, children operate somewhat as scientists do, collecting evidence about the world, evaluating that evidence, and constructing theories to explain what they have observed. Cognitive change occurs when enough counterevidence has built up against an existing theory that an alternative theory becomes more attractive to the child. To spell out the full implications of saying that young children literally have *theories* about the world requires both conceptual inquiry into what can, and should, be meant by 'theory' in this context, as well as a factual inquiry into the nature of theories and theoretical change. The factual inquiry may be partly empirical, grounded in observations about theories and theoretical change that have been revealed by the history of science, for example, and it may be partly based on intuitions about the nature of theories. With broad factual knowledge about theories, coupled with conceptual work on how the term 'theory' ought to be employed for the purposes at hand, the claims of the theory theory can be evaluated much more clearly. For example, we might argue that, intuitively, if children's beliefs are theoretical they should also be revisable in the light of new evidence, and we might then look for evidence of such revisability.

A particularly common theoretical use of intuitions in science involves a kind of analogical reasoning. Science often proceeds by the use of analogy. Electric current is likened to the flow of water through pipes, the atom is likened to the solar system, the transmission of light is likened to the movement of a wave through water. In each of these cases, there is, to use Mary Hesse's (1966) terminology, both a "positive analogy"—respects in which the two systems are thought to be similar—and a "negative analogy"—respects in which the two systems are thought to differ. The use of analogies in these cases both helps us to understand an unfamiliar phenomenon in terms of a more familiar one and suggests areas for testing and research, to see how far the positive analogy can be pushed. If light is really like a wave, then it should do such-and-such in such-and-such circumstances that we have not yet tested. The results of such tests occasionally are stunningly successful. Fresnel's wave model of light predicted the appearance of a bright spot in the center of the shadow of a small circular disk when a narrow beam of light was shone upon the disk, and much to the surprise of many scientists at that time, when the experiment was conducted the bright

spot was found. Of course in the twentieth century, the wave analogy for light has produced some equally stunning predictive failures.

To return to the example of the “theory theory,” the more we know about theories, and the more precisely we can articulate our concept of a theory, the better sense we can make of the use of the term ‘theory’ in empirical psychology, whether that use is meant literally or analogically. The use of ‘theory’ in the theory theory is, in fact, a good example of a case where literal and analogical uses shade into each other. Should the theory theory be seen as regarding children as *literally* possessed of theories, or should it see their cognitive changes simply as *analogous* to scientific theory change? A theory theorist’s answer to this question may be largely a matter of taste.

Some of the most significant advances in cognitive psychology and cognitive science have taken intuitive, commonsense, or “folk” psychology as their base and used it analogically, modifying and expanding it to apply to new kinds of phenomena, much as our language and knowledge about waves was expanded to apply to the domain of optics. The best functionalist cognitive psychology, in this vein, takes folk psychological notions like “inferring” or “reasoning” or “following a rule” or “believing” or “desiring” and applies them in cases where no folk psychologist ever would. Cognitive psychologists have used the vocabulary of belief, inference, rule-following, and achieving goals to explain phenomena like perceiving the moon on the horizon, uttering a grammatical sentence, or automatically and unconsciously reaching for a close object. Indeed, it might be argued that the very idea of computation, the centerpiece of cognitive science, is itself derived from folk psychology. Alan Turing (1950) begins his classical discussion of computation, after all, with an account of what it is like to do a particular kind of conscious problem solving, an account that would accord with anyone’s folk psychological intuitions.

Rather than assimilating the folk psychological notions wholly without modification, however, Turing, Chomsky, and other modern theorists of cognitive science expanded, modified, and formalized these notions, applied them to new domains, and tested them with new kinds of evidence. These extensions have notoriously violated the intuitions of folk psychology itself. Indeed, it is precisely this fact that makes the extensions and applications interesting. Claiming that we compute when we add a column of numbers or follow rules when we play Monopoly is uninformative at best. Claiming that we compute when we perceive or follow rules when we talk has, in contrast, been very scientifically productive. Analogical uses like this have a way, if they are successful, of becoming literalized over time. We came to say, literally, that light was a kind of wave (though twentieth-century science presented some complications here). We now say that, literally, electrons have orbits. At some point—perhaps now—we may also want to say, literally, that the visual system computes.

One Prevalent Misuse of Intuition

Because the cognitive sciences have employed our intuitions in the diverse ways we have described above, they have profited enormously from the philosophical

work of articulating intuitions and drawing out the implications of our intuitive views. What is it, then, that leads to the sort of dialogue with which we started this essay? What is wrong with Phil? Our answer, only partly ironic, is that the poor kid is the victim of a category mistake. Within philosophy of mind, there has been a tradition of either accidentally confusing different uses of intuition, or deliberately treating them as identical, and consequently drawing inappropriate conclusions. An unspoken assumption of much argumentation in the philosophy of mind has been that to articulate our folk psychological intuitions, our ordinary concepts of belief, truth, meaning, and so forth, is itself sufficient to give a theoretical account of what belief, truth, meaning, and so forth, actually are. We believe that this assumption rests on an inadequate understanding of the nature of intuition and its appropriate applications, and that it results in errors of the sort that the case of Phil dramatically illustrates. The error is more vivid in Phil's case only because the intuitions to which he appeals are so clearly (to us) out of step with the world.

We will briefly discuss three notable examples of what we take to be this sort of misuse of intuition in philosophy. We will focus, in particular, on cases in which developmental psychology provides relevant evidence. The first example occurs in the long and fevered debate about externalism in semantics, prompted by Hilary Putnam's (1975) and Saul Kripke's (1972) philosophical work. These authors articulate in an interesting and perceptive way an unexpected and surprising intuition we ordinarily have about meaning (at least many philosophers seem to have found this intuition surprising). The intuition is that meaning is, in Putnam's words, not in the head. We have the intuition that what a word means, for a person, depends not only on what is going on internally with that person, but also on facts about the world external to the person. This is the point of Putnam's Twin Earth example, described above. That we have this intuition turns out to be an important and interesting fact about us, one that a scientific psychology would want to explain. The intuitions described by Putnam, Kripke, and others after them are not just exotic philosophical intuitions—Frank Keil (1989) and Susan Gelman (Gelman and Wellman 1991) have shown that they may be shared even by fairly young children, and cognitive psychologists like Murphy and Medin (1985) and Rips (1989) have shown that they are shared by ordinary adults and play an important role in adult action and judgment. This intuition is an important piece of data for empirical psychology; Keil and the rest are treating the Twin Earth intuitions evidentially.

Moreover, Putnam's and Kripke's arguments alert us to the fact that our standard internalist psychological theories of meaning do not exhaust the theoretical possibilities. We might apply them to generate new hypotheses about the nature of meaning. Psychologists interested in the internal structure of the individual's mind may naturally be drawn toward an internalist account of meaning that sees the meanings of a subject's words as dependent solely on the internal states of that subject. However, science has often profited enormously from relational, supra-individual concepts that ascribe properties to individuals not solely on the basis of facts internal to that individual. Explanatory notions like "fitness" in biology are inherently relational and supra-individual in this way. Perhaps it will turn out that the notion of semantics we need to do explanatory work in scien-

tific psychology will be similarly relational. So, in addition to being an interesting piece of psychological data, this intuition provides an interesting hypothesis about the nature of meaning.

We should not confuse these two applications of the Twin Earth intuitions however. These intuitive hypotheses about meaning need not be scientifically viable in the long run. That is, the fact that our ordinary understanding of the mind presupposes an externalist, supra-individual semantics says very little about whether an externalist semantics is the best choice for a theory of meaning. It may be the case that a relational, supra-individual conception of meaning does the best job in explaining meaning. But then again, it may not be. The fact that the ordinary conception of meaning within our daily experience has this character provides only rudimentary evidence about what notion of meaning will work best for psychologists and others who wish to talk about meaning in a scientifically informed way—just as our ordinary intuitions about physics provide only a rudimentary starting point for the creation of physical laws. These intuitions may have played a crucial initial role for Galileo, but most of them did not survive in the end.

Empirical evidence about the nature of belief is irrelevant to Phil's observation that his three-year-old concept of a belief is a concept of something that cannot be false. Psyche's observations cannot undermine this claim of Phil's, and to the extent that Phil is only articulating his current concept of the mind, he is safe from cognitive science. Indeed those intuitions will be an important source of data for developmental psychologists. However, when Phil goes on to infer that his intuitions about the mind, treated as hypotheses about human cognition, are similarly untouchable by empirical observation, he has committed an error. Any philosopher who claims that our intuitions about Twin Earth reveal some scientifically unassailable fact about the human ability to mean things by words—rather than simply revealing something about what we *mean* by 'meaning'—makes an error similar to Phil's.

A second example of the error is in John Searle's critique of cognitive science (Searle 1990, 1992). Searle observes that in our ordinary intuitive understanding, states with intrinsic propositional or intentional content like desire and belief are bound up with conscious experience, or phenomenology, in a particular way. Indeed, in our everyday experience mental states and conscious states almost always co-occur: The creatures or systems that occupy states whose causal, functional roles are the same as those of belief, desire, and other states with intrinsic intentional content have always been the same systems that had conscious phenomenology. Consequently, the intuitive conception we have of intrinsic intentional states presupposes the co-existence of these two factors. Now, suppose it turns out, as it seems it might, that there are systems that, behaviorally speaking, occupy the causal roles associated with intrinsic intentional states—they act as though they have, for example, beliefs—but do not have conscious phenomenology. In that case the intuition is foiled, at least if it is construed as a hypothesis about the necessary relation between certain causal, functional facts and phenomenology. Searle acknowledges the possibility of such systems, but he insists that such systems cannot be said *really* to have intrinsic intentionality.

Now, in fact, cognitive scientists have done a lot of interesting work on the

mind that depends on ascribing states with intrinsic intentional content to sub-systems of the mind that generate no phenomenology, or at least no phenomenology of the sort that is intuitively associated with the intentions that are being ascribed. Noam Chomsky's work on grammar (Chomsky 1980) and David Marr's work on vision (Marr 1982) are two prominent examples. In general, computational theories of the mind depend on describing the mind as engaged in all kinds of computational processes (with intrinsic intentional content) unconnected or only remotely connected to conscious phenomenology.

So what is a scientist to do? If we can better understand, explain, and predict scientifically what is going on in such systems when we treat them as though they had intrinsic intentional content, then it would be a mistake to forbid doing so, as Searle seems to want. Everyday intuition does not and should not have authority over how terms and concepts are to be applied and extended in the sciences; if science benefits from declaring that intentional states "really" need have no phenomenal accompaniment, philosophy ought not prevent such a conceptual change. This is not to say that Searle or any other philosophical skeptic about this aspect of cognitive science has no recourse: He can argue that science does not benefit from such an analogical extension of our intuition, and he can caution against certain inferences that may depend upon aspects of the old conception that have been cast aside. But it is illegitimate to put an analysis of our intuitive concepts to work as a block to the movement of science in one direction or another, as though it revealed some independent truth that science was powerless to overthrow. A potential compromise position would allow cognitive scientists to employ the vocabulary of intrinsic intentionality, all the while insisting that the use is analogical (or in Searle's terminology, "as-if"). In fact, some cognitive scientists may have something like this in mind—although it would certainly be understandable were they to tire of insisting that the use is analogical and decide to broaden the literal meaning of these terms to cover the cases at hand.

If the intuitions about intentionality to which Searle is appealing are simply treated as data, then we can vest some level of confidence in the truths *about our concepts* that we so discover, but if we treat our intuitions about intentionality as hypotheses about how the world really is independent of our concepts—that is, if we treat them as hypotheses about the mind—then they have at best a rudimentary, preliminary, and *prima facie* warrant. One way, though perhaps not the only way, of interpreting what we take to be Searle's error in his treatment of cognitive science, is that he employs his folk psychological intuitions hypothetically to reveal something about the world beyond our concepts, but he sees them as having a degree of warrant that is justified only when they are employed to reveal the content of our present concepts. A similar error is made by those who would, in an argument that is in some ways the mirror image of Searle's, take the connection between functional role and phenomenology in ordinary intuition as some kind of proof that conscious phenomenology will necessarily emerge if ever we create a robot with the right functional relations between its inputs and its behavioral outputs.

A third example, which has been discussed at greater length in Gopnik (1993), is the question of the nature of first-person authority about our mental

states, our conviction about the accuracy of our own reports of our beliefs, desires, moods, and so forth. A wide variety of philosophers have defended first-person authority by pointing to the intuitive links between first-person authority and other deeply embedded aspects of our commonsense psychology. Thus, Davidson (1987), for example, suggests that first-person authority is intimately bound up with and presupposed by certain facts about the interpretation of language, while Shoemaker (1988) more simply, suggests that first-person authority is bound up with our very concept of belief. Even if these philosophers are entirely correct in the ways they have spelled out our concepts and our intuitions about the mind—even if it is impossible for us to conceive of belief without conceiving of it as known to us first-person, or to conceive of ourselves as responsible or interpretable agents without first-person authority, this does not settle the question of whether we actually do have first-person authority. The fact that we are so convinced that we do have first-person authority is an interesting and important fact for psychological theory to explain. And perhaps psychological research will in fact demonstrate that our intuitions of first-person authority are not in error—there may be links, for example, between being in a certain brain or functional state and having certain subjective experiences, such that having those experiences provides one with a special warrant for claiming that one is in those states (see Goldman 1993b for such an argument). But the mere fact that our everyday folk psychological intuitions are congruent with such a picture does not imply that that picture is correct as a deeper theoretical account of our psychological structure.

Why is this kind of confusion about the appropriate application and warrant of intuitive judgments so widespread in philosophy of mind? To sound like Phil, here is a diagnosis and a suggested therapy. The diagnosis is that the confusion is the result of the relatively close relation, at this stage of the game, between the structure of our folk psychological everyday conceptions of the mind and our scientific conceptions, and the very fertility of our scientific borrowing of folk notions. Our scientific accounts, quite properly and productively, take off from the folk account. Our ordinary psychological intuitions are the first and best source of hypotheses about how the mind works. As long as the concepts and structures of the two accounts are relatively similar, they are easily confused.

If (or when) psychological theories are formulated in terms that are incongruent with intuitive psychology, the independence of the two becomes clearer. For example, we know that much of our low-level visual perception can be explained as a system that performs a Fourier analysis of information in the retina. We could imagine making a Searlean objection that perception does not feel like Fourier analysis. That objection, though, would seem patently to miss the point. The reason is that there is nothing that we intuitively think Fourier analysis feels like. The concept plays no role at all in our intuitive psychology. In contrast, when we say that perception involves inference, an elision is much more likely between this theoretical, analogical extension of our folk psychological concept and its application in folk psychology itself.

It is interesting to note that, historically, philosophy has abandoned the project of articulating our physical intuitions about the world or using those intuitions as evidence about the underlying structure of the world. Perhaps the last

great effort in this direction was made by Kant and his followers in the eighteenth and nineteenth centuries. Kant (at least as he is often interpreted) supposed that an examination of the necessary presuppositions of our physical intuitions would reveal something about the structure of the physical world, just as contemporary philosophers of mind suppose that an examination of the necessary presuppositions of our psychological intuitions will reveal something about the structure of the mind.

The abandonment of the Kantian project was the direct result of the great conceptual and empirical progress in physics over the last two hundred years. The advances of physics and its eventual deep departure from our ordinary physical intuition made the distinction between the conceptual structure of our “folk physics” and the conceptual structure of theories in scientific physics painfully clear. The devastating effect of progress in physics on the Kantian project was so absolute that it is difficult even to understand the appeal of the project now. There is no contemporary philosophy of body or of space and time that does not begin with an intimate knowledge of empirically justified scientific research.

Similar progress in psychology might well lead to a similar result and to similar changes in what are seen as appropriate enterprises for philosophy. This argument has, of course, been made at some length by Paul Churchland (1981) and Stephen Stich (1983) among others. But, it may be countered, this is unfairly eschatological. We have no way of knowing, now, that a scientific psychology will in fact have the success of a scientific physics or depart from ordinary intuition in the same way. It seems unconvincing to counter claims that intuition is a genuine guide to psychological fact by making reference to a possible future psychology that will consign those intuitions to the dustbin of history. Psychology is different from physics, and perhaps indeed the nature of those differences is such that no scientific psychology that is widely different from folk psychology will ever be possible.

Fortunately, we need not wait for a psychological Einstein to be convinced of the contingency and instability of even our deep-seated intuitive psychological beliefs and experience. We have no crystal ball that lets us view the future of our psychology, but we do know quite a lot about its past. By looking at development, an almost completely neglected source in philosophy of mind, we can make much the same point as Churchland and Stich without the eschatology. We can be confident that our currently constituted folk intuitions are not the last word about the mind because they were not the first word. The best therapy for the confusions outlined in this section, we would argue, as indeed for many other adult ills, is to spend more time with children.

What Children Tell Us About Intuition

In some respects the project of cognitive developmental psychology is more like the philosophical project of elaborating and articulating commonsense intuitions than it is like standard, grown-up cognitive psychology. Just as one aim of philosophy is to articulate certain sorts of adult intuition, one aim of developmental psychologists is to characterize the intuitive beliefs of children. Unlike philoso-

phers, however, the developmentalists need to use more systematic means than introspection to accomplish this end, hence their laboratories, research budgets, and mastery of ANOVAs. In using experimental techniques to discover how minds intuitively construe the world, development joins hands with such diverse aspects of psychology as clinical cognitive neuroscience, ethology, and cognitive anthropology. By looking at the intuitions of these diverse populations, we can evaluate and understand the nature of the intuitions, provided by philosophy, of healthy, adult, Western, human beings.

The developmental data suggest that children's intuitions about quite standard psychological (and physical) matters are often quite different from those of adults. There is by now an extensive literature in cognitive development and theory of mind that charts these differences. We will briefly mention three examples that parallel the three examples of the philosophical misuse of intuition described above.

First, we now have extensive studies of children's understanding of "natural kinds" in their folk semantics and ontology (Keil 1989; Gelman and Wellman 1991; Carey 1985). The studies suggest that children's semantic intuitions are both similar to and different from the intuitions of adults in systematic ways. Even very young children, as young as two-and-a-half or three, seem sometimes to assume that words refer to the underlying causal powers of objects rather than to their superficial perceptual features (Gelman and Wellman 1991). On the other hand, children seem to be less sensitive to the constancy of causal powers over superficial transformations. Keil, for example, asked children a series of "Putnamesque" questions about the fate of a cat that was perceptually transformed to look like a skunk (it was painted black with a white stripe, it had a smelly bag added, etc.). Children seven or eight years old showed a pattern of intuitions like the Putnam intuitions—they said the cat would still be a cat and should still be called 'cat'. Equally significant, however, children five or six years old did not show this pattern of intuitions. Their intuitions told them that the cat would indeed become a skunk and be called 'skunk' under these conditions. Using another Putnamesque intuition pump yielded still another pattern. Children eight or nine years old said that if scientists discovered that skunks had the same internal structure as cats, they would be cats and should be called cats, but the younger children who showed "natural kind" intuitions about transformations did not have similar intuitions about scientific discovery, although they understood the nature of the scientists' work. In short, some aspects of the Putnam intuitions, such as attention to underlying causal powers, seem to be in place very early, while other intuitions emerge gradually with increasing knowledge and experience.

A second example concerns children's intuitions about the relations between intentional states and conscious phenomenology. In a long and elegant series of studies, John Flavell and his colleagues (Flavell, Green, and Flavell 1995) have explored children's intuitions about conscious experience. Children, like adults, gave every evidence of knowing and recognizing what conscious experience was. Earlier studies (for example, Estes, Wellman, and Woolley 1989) found that even three-year-old children could discriminate with some sophistication between the thought or image of a hot dog and an actual hot dog. They reported, for example,

that the thought was internal, private, and subject to change by the will alone, while the real hot dog was not. However, in Flavell's studies, children seemed intuitively to associate conscious phenomenology with a different set of stimuli and behavioral patterns than adults did. For example, children reported that someone who was wide awake but was simply looking at a blank wall would have no conscious phenomenology, no stream of thoughts, impressions, images, and so forth. More strikingly, they reported that they themselves had no conscious phenomenology in similar conditions, despite the availability of a wide range of words with which to report such experience. In one experiment, Flavell et al. rang a bell at regular one-minute intervals. On the last trial, the bell only rang after two minutes. When they asked the children about their conscious phenomenology in the delay period, children said that they had not been thinking about or expecting the bell, and that no thoughts of the bell had entered their heads. In contrast, children sometimes ascribed phenomenology in conditions that adults would not. They reported that people did have conscious phenomenology in dreamless sleep and that therefore one could, for example, decide to turn over in bed while deeply asleep. Again, even relatively old children, in this case five- or six-year-olds, seemed to have quite different intuitions about the relation between phenomenology and function than adults do. They certainly had different intuitions than John Searle does.

The last example comes from the first author of this chapter's work on children's reports of their own past mental states and the mental states of others (Gopnik and Astington 1988; Gopnik and Slaughter 1991). As we outlined at the beginning of this chapter, three-year-old children appear to think that beliefs, their own beliefs and those of others, are necessarily true, and they misreport their own immediately past beliefs. In contrast, these three-year-old children seem to have much the same intuitions as adults about simple perceptions. They report that their current perceptions may differ from those they had in the past and may differ from those of other people. Three-year-olds are also better at reporting changes in desires or in the appearances of objects than they are in reporting beliefs. Moreover, recent evidence from our lab suggests that even younger children, twenty-four to thirty-six-month-olds, do not make the same predictions about perception that adults and three-year-olds do. In short, there seems to be a gradually developing sequence of folk psychological intuitions in children, just as there is a gradually developing sequence of folk semantic intuitions in the work of Keil and Gelman, and a gradually developing sequence of intuitions about the occasions of phenomenal experience in Flavell's work. As these folk psychological intuitions develop, children become more reliable sources in reporting their own mental states, including their present and recently past phenomenology and their present and recently past beliefs and desires.

It is important, but difficult, to remember that when we talk about "children" we are talking about our past selves. Hilary Putnam, John Searle, Sydney Shoemaker, and the rest of us once, not so long ago, had very different intuitions about the nature of meaning, phenomenology, and belief than we do now. Our intuitions are not constant; they have changed quite radically in the past.

There are two possible explanations for these differences and changes. One possibility is that the underlying psychological structure of children, even seven-

or eight-year-old children, is radically different from our own, and so their intuitions are correct about three-year-old minds but not adult minds. This, in spite of the fact that three-year-olds talk, behave, interact, and function in very many ways just as we do. Were it not for the research of developmental psychologists we would not, in fact, have any idea that these children had different intuitions than we do. In order to preserve the idea that intuition is an accurate guide to underlying psychological structure, we would be forced to this radical and unpalatable conclusion.

The alternative and more plausible explanation, the one favored in developmental psychology, is that children have quite similar underlying psychological structures to our own, and that a theory of their semantics, consciousness, and belief would be similar to a theory of adult semantics, consciousness, and belief. However, their conception of that psychological structure, their knowledge about it, their folk psychology, is quite different from our own. The most likely explanation of this difference is quite straightforward: Children know less about the mind than adults do, and they learn more as they grow older. And, quite naturally, as they learn more about the mind, their conceptions about the mind, and so their intuitions, change.

But then, of course, the eschatological possibility described by Churchland and Stich becomes not only possible but likely. If our intuitions about the mind were mistaken in the past, and the error of these intuitions was revealed as our knowledge about the mind grew, then surely our intuitions about the mind now, as imperfect adults, can be mistaken in ways that will be revealed as our knowledge about the mind grows. This may already be happening, as twentieth-century adult intuition changes to accommodate newly popular concepts, like that of the unconscious. The four-year-old Phil looking back on his earlier self would, we hope, be struck by the folly of his earlier errors. Perhaps also he would be preserved from making similar errors in the future.

Notes

The preparation of this paper was supported by NSF grant DBS9213959 to the first author. We are grateful to John Campbell, Clark Glymour, and Pauline Price for helpful discussion and comments, and to the organizers and participants in the Notre Dame conference.

1. Apparently McDowell (1994) had found its way on to the picturebook shelf.