3 Folk Psychology and Tacit Theories: A Correspondence between Frank Jackson, and Steve Stich and Kelby Mason

Frank Jackson, Kelby Mason, and Steve Stich

K&S[1]: Letter from Kelby Mason and Steve Stich [1] to Frank Jackson, February 20, 2005

Dear Frank,

It's been a while since our paths have crossed. I hope all's well in your part of the world, and that you and your family are flourishing.

A few months back, Robert Nola contacted me to tell me that the "Canberra Plan" volume had found a publisher and to ask whether I was still interested in trying my hand at a correspondence with you—of the sort we had discussed one lovely morning on my last trip to Sydney. I said that I was and I asked whether I could do my part collaboratively with Kelby Mason. Robert said that was fine with him.

If you're still interested in pursuing the idea, let me suggest a few vague ground rules. Our initial idea in Sydney was to exchange a few letters to try to understand where we differed about the nature of folk psychology and the relevance of various putative facts about cross-cultural variation in intuition. That's still my interest (and Kelby's). No doubt we'll have plenty to debate, but I see the project less as a debate and more as an opportunity to clarify your position and mine. To that end, I suggest that Kelby and I begin by posing a few questions to which you can reply. We'll react to your replies and respond to any questions you might raise. We can keep the process going as long as it seems fruitful. So much for our suggestions on ground rules. If you're not comfortable with any of this, we're certainly open to alternative proposals.

So now let's turn to the questions. If memory serves, in the conversation which first gave rise to the idea of this exchange, you said that you did not think of folk psychology as something in the head. This took me aback, since I had always assumed that you thought of folk psychology as an internalized theory, more or less as I did. It suggested that at least some of our disagreements about folk psychology are not real disagreements at all, since we may be talking about different things. On and off, since then, I've tried to come up with a charitable interpretation of your brief comments in Sydney. But I haven't been very successful. So let me start by asking:

Q1 Do you still think that something that you'd want to call "folk psychology" is not in the head? If so, can you explain what it is?

Since it hardly seems fair to just drop these questions on you and let you do all the work, let me offer a bit of context. In a paper I wrote with Ian Ravenscroft a bit over a decade ago, we distinguished six different answers to the question: "What is folk psychology?"[1] Four of the answers we offered maintain that folk psychology is "an internally represented 'knowledge structure' used by the cognitive mechanism underlying our folk psychological capacities" (Stich and Ravenscroft 1996, 132), where "our folk psychological capacities" include our capacity to attribute mental states to ourselves and others and our ability to predict and explain people's behavior (ibid., 124–126). What distinguishes the four answers is how the information is stored. Since you maintain that folk psychology is not in the head, I assume that *none* of these accounts is very close to what you have in mind. Is that right?

The two remaining accounts Ravenscroft and I offered are what we called "external" accounts of folk psychology. The first of these is simply "the set of folk psychological platitudes that people readily recognize and assent to." The second is "a theory that systematizes the folk psychological platitudes in a perspicuous way" (ibid., 132). There is a straightforward sense in which neither of these is "in the head," which is why Ian and I called them "external accounts." But on my (unfortunately very dim) memory of what you said when I asked for more information about your conception of folk psychology, I don't think either of these accounts is at all close to what you had in mind either. So your conception of folk psychology is not on the chart that Ravenscroft and I constructed. Nor, I assume, do you have in mind the sort of account of folk psychology that simulation theorists urge, and that Ian and I discuss briefly in section 5 of our paper. So there are seven things you *don't* mean. And at this point Kelby and I need some help. If folk psychology, for you, is not any of these, what is it?

Of course, folk psychology is only one sort of folk theory that philosophers are interested in, so our second question is simply a request for clarification:

Q2 Does what goes for folk psychology go more generally for folk theories? That is, is the story you tell about the nature of folk psychology the same sort of story you'd tell about, say, folk epistemology, folk ontology, or folk metaphysics?

If it is, then any disagreement we have here is just one instance of a broader disagreement about the nature of folk theories. So, in our discussion, we should keep in mind the broader picture, although folk psychology is probably a good test case for us to focus on.

There are two other important questions that Kelby and I would like to discuss at some point. Since your answers to them will depend on your answers to the first two questions, we don't expect you to address them just yet. But we do want to flag them now for later discussion. On the two "external" accounts of folk psychology mentioned above, platitudes or intuitively obvious claims play a central role in characterizing folk psychology. You have also stressed the role of intuitions in conceptual analysis in other domains. And, as you know, along with a gaggle of former students, I have been exploring intergroup variation in various sorts of philosophical intuitions.[2] The first of our two additional questions is:

Q3 What do you think would follow if there really is *substantial* and *systematic* variation in intuitions between different groups (e.g., cultural, racial, age, or gender groups)?

Our second additional question is this:

Q4 How do we decide, in the course of conceptual analysis, which intuitions are *not* a good guide to folk theories (and therefore should be ignored)?

The Chomskian tradition, for instance, has an answer to this question which seems well motivated, although often hard to apply: some grammatical intuitions reflect (or accord with) the rules that speakers have internalized; others are "performance errors" which result from memory limitations, failures of attention, interference from other mental systems, etc. Now, it would be hard enough to extend this Chomskian answer to folk theories in general, even if you did think that folk theories are "in the head." Since you don't think they're "in the head," it's even harder to see what kind of alternative answer you could give. As I say, we don't expect an answer to Q3 or Q4 yet, but we would like to return to them eventually.

OK. That should be enough to get the ball rolling. Kelby and I look forward to reading your responses.

With warmest regards,
Steve and Kelby

FJ[1]: Letter from Frank Jackson [1] to Steve Stich and Kelby Mason, April 27, 2005

Dear Steve and Kelby,
Here's my belated reply to your letter of 20 February. I'll try and say as clearly as I can where I stand on folk psychology and how this connects with your questions.

1 Is Folk Psychology in the Head?

As I use the phrase 'folk psychology' it stands for a certain theory about what the world is like. The theory says that there are states that stand in such and such relations to each other and to events in our world. I hold that very many people believe this theory, which is why I think it is right to call it a folk theory. In the same way I think that many people hold the theory that, as a rule, unsupported bodies fall. In consequence it is right to call this a folk theory. It would not be right to call quantum physics a folk theory because very few people hold it (lots hold that it is true without knowing what it is that is true).

Is the theory that, as a rule, unsupported bodies fall 'in the head'? Not in any natural sense. I think of theories as individuated by their contents—how things have to be for them to be true—and contents aren't in the head—as Bob Stalnaker often remarks.

Of course, there is the question of what makes it true that some person holds some theory, and I do think that that is in the head (as a matter of contingent fact). I think, that is, that it is how a person is inside that settles what they believe, although typically what they believe concerns how things are outside them. In the case of folk psychology, I think the content—what they believe—is partly about how things are inside persons who hold the theory in the sense that the states they hold stand in such and such relations include some that they also hold are inside them, partly about the insides of other people, and partly about how surroundings interact with people's insides generally.

So what I am saying is that we need to distinguish the content of a theory from the holding of that theory when we ask if the theory is in the head. The content isn't inside the head but what makes it the case that they hold the theory is inside the head. This applies to theories in general; it is not a thesis about folk psychology or folk theories in particular. The content of the Big Bang theory isn't in the head but what makes it true that someone holds it is in their head. (This does not mean that the content is an intrinsic property of their head state. It is inter alia how their head state interacts with the environment that makes it the case that they have the belief with the content that it all started with a big bang.)

Here is how this distinction affects the discussion in your letter. You mention in K&S[1] the view that folk psychology "is an internally represented 'knowledge structure.'" From my perspective, this could not possibly be a good account of the content of folk psychology, any more than it could be a good account of the content of the folk theory that, as a rule, unsupported bodies fall. What might be true is that what makes it true that a subject believes in folk psychology is that he or she has such an internally represented knowledge structure.

Or consider one of the 'external' accounts of folk psychology you mention in K&S[1]: "The two remaining accounts Ravenscroft and I offered are what we called 'external' accounts of folk psychology. The first of these is simply 'the set of folk psychological platitudes that people readily recognize and assent to.'" This is a view about the content of folk psychology to the effect that its content is one and the same as that of the conjunction of the platitudes as specified (not a view I hold, as it happens, but one Lewis once held, more or less). It is not in competition with various accounts of what it takes to hold the theory—that is, it is not in competition with various views about what it takes to hold a theory with the same content as the conjunction of the platitudes.

2 The Subpersonal versus Personal Level Question

On discussing the 'in the head' issue with Daniel Stoljar, he suggested that the question you may have in mind is the personal versus subpersonal question (and now I come to think of it, this may have been the question under discussion in Sydney).

Defenders of folk psychology often say that it is an implicit theory. I do. But what I mean by this is one, but only one, of the things sometimes meant when it is said that we have an implicit theory of grammar. There is an implicit theory that drives our classifications of sentences in languages we have mastered into the set of the acceptable and the set of the nonacceptable sentences. This is the theory we make explicit by interrogating our intuitive classifications and which, when extracted and recorded in words, makes its way into grammar books as an explicit theory of grammar. That's how grammar books get written.

What I do not mean is the sense of implicit theory in which it is said that we have an implicit theory at the *subpersonal* level. Our ability to classify sentences into the grammatical and the nongrammatical must have an explanation at the subpersonal level. In this regard it is like our ability to locate sounds. The explanation for this is, in part and roughly, that our brains latch onto the relevant out-of-phase effects that occur in the inputs to our ears. I do not know if anyone knows the corresponding explanation in the grammar case but there must be one, and that is what some have in mind when they talk of our implicit theory of grammar. If we call the first the personal-level implicit theory, and the second the subpersonal-level implicit theory, what I am saying is that we have an implicit theory at the personal level and at the subpersonal level, both in the case of grammar and in the case of folk psychology.

Why do I think we have a personal-level implicit theory in the case of grammar? Because no brain science was needed to write grammar books. All the same a fair bit of work was required, so it wasn't explicit in any obvious sense. The task was to put into finite strings of words the pattern we recognize at the personal level and that was hard work.

Why do I think that we have a personal-level implicit theory in the case of folk psychology? It isn't explicit—if it were, how come there's so much argument? But if all we had were an implicit theory at the subpersonal level, we would not know what we were saying when we said that someone was in pain. When I say that a sound is located at *L*, I am not saying that the out-of-phase effects at my ears are thus and so. The content is not given by the subpersonal level account of how our brains do the locating job. Ditto for folk psychology, say I. The content is not given by the way the brain makes patterned sense of it all at the subpersonal level.

3 Does What I Say for Folk Psychology Apply to Folk Theories across the Board?

My view about folk psychology is a view about its content. The content of other folk theories will be quite different of course. So the answer is no if the question is one about similarity of content. If the question is, Do I hold that all folk theories are to the effect that various states stand in such and such relations, with the difference between folk theories being in the specified relations, the answer is also no. I think the folk hold that some things are round but that's not a matter of relations between states.

4 What Follows if There Are Substantial and Systematic Variation in Intuitions Between Different Groups?

It would follow that there was a difference in concept (unless they were confused—more on this in later exchanges). A live example in my view is Twin Earth. There may well be subjects who insist that XYZ counts as water. In that case their concept of water differs from ours.

5 Which Intuitions Are a Good Guide?

I have no problem with the Chomskian answer you mention. I'd only add that there are two things one can mean by getting the answer wrong in the grammar case. You can mean getting the answer wrong in the sense of not conforming to the norm, or you can mean getting the answer wrong in the sense of not conforming to one's own concept. The latter is less common than the former and it is in the latter case that performance errors most obviously enter the picture.

I hope all this helps and many apologies for being so slow.

All the best,
Frank

K&S[2]: Letter from Kelby Mason and Steve Stich, May 5, 2005

Dear Frank,

Many thanks for your letter of April 27. It did an excellent job of answering some of our questions, and raising some important new ones—all with admirable clarity and brevity. What we'll do in this letter is (1) summarize our reading of your answer to one question (viz., "Is folk psychology in the head?") and (2) explain why we think your comments about the personal and subpersonal levels and implicit theories raise a cluster of new questions. We suspect these new questions may underlie whatever disagreement we have on these matters.

First, however, let us note a very insignificant point that became clear to us in reading your letter. As you use the word 'theory', a theory can be a very small packet of information (or misinformation). For example, you write: "I think that many people hold the theory that, as a rule, unsupported bodies fall." Other writers pack more into the notion of a theory, insisting that a theory must be a fairly complex, interrelated cluster of propositions, and some add that a theory must posit its own set of "theoretical entities."[3] There is, as best we can tell, exactly nothing of serious philosophical interest at stake here. We mention it only because one of us (Steve) has, in the past, been confused by comments like the one displayed above, and we suspect that other readers may have had much the same ". . . but that's not a *theory* at all!" reaction.

OK. Now let's get on to more substantive matters.

1 Is Folk Psychology in the Head?

Though we had both been puzzled, in the past, by your insistence that folk psychology [FP] is not in the head, your letter has—if we are reading it properly—completely eliminated that puzzlement. As we read your letter, the crucial passage is: "Is the theory that, as a rule, unsupported bodies fall 'in the head'? Not in any natural sense. I think of theories as individuated by their contents—how things have to be for them to be true—and contents aren't in the head—as Bob Stalnaker often remarks."

What is important, here, is that when you talk about a theory, what you have in mind is the content of the theory (or, to be a bit fussy, something that is individuated by its content), and content "ain't in the head." So what's doing most of the work, here, is content externalism (aka meaning externalism), which, of course, comes in many varieties. Whether or not meaning externalism is true is a contentious matter. And one of us (Steve) has notoriously argued that the debate is deeply flawed since the participants have not said what counts as getting a theory of content right.[4] But, fortunately, we need not debate any of these issues here. If, in claiming that FP is not in the head, you were claiming something akin to what many other philosophers have meant when they say that beliefs ain't in the head, then we are no longer puzzled by your claim.—Progress!

2 Personal versus Subpersonal

In your letter, you suggest that "the question [we] may have in mind is the personal versus subpersonal question." We suspect you are right about this. The distinction seems to play a crucial role in your thinking about FP and other meaning-anchoring folk theories. And, to put our cards on the table, we doubt there is a distinction to be drawn between personal and subpersonal theories (or levels, or explanations) that will do the work you need it to do.

We can think of *some* ways of drawing the personal/subpersonal distinction which are both reasonably clear and potentially useful for some purposes. But we don't think the distinction drawn in these ways is the distinction you have in mind. For example, some people seem to use the "personal/subpersonal" distinction as a way of distinguishing information that is consciously (or introspectively) accessible to a person from information that is not. So, for example, if asked whether the tone she hears is coming from the left or the right, Sally will typically be able to introspect and answer with confidence. But asked whether the tone she hears in her left ear is out of phase with the tone she hears in her right ear, Sally will say that she has no idea. Introspection is no help. We don't think that the *accessible-via-introspection/not accessible-via-introspection* distinction can be identified with the personal/subpersonal distinction that you have in mind, however, since, as you note in your letter, in both the grammar case and the FP case, people can't just introspect to determine whether a particular claim is part of their "implicit" personal-level theory.

Before going on, let us raise a pair of related questions, to be sure we're all "still on the same page," as they say.

Q5 In your letter you use both 'explicit' and 'implicit' on a number of occasions. As we read you, these *are*, near enough, the equivalent of our '(readily) accessible via introspection' and 'not (readily) accessible via introspection'. Is that correct?

Q6 Are we right that you do *not* consider the personal/subpersonal distinction to be the same as the accessible-via-introspection/not accessible-via-introspection distinction?

If the answers to Q5 and Q6 are *yes*, then presumably you think there are two kinds of implicit theory that might be attributed to someone: an implicit theory at the personal level and an implicit theory at the subpersonal level. However, this is the distinction that we do not understand.

To say why, let's consider the case of grammar. Let's assume that one of the goals of grammar is to specify the set of sentences that a speaker of a language finds acceptable from the set of sentences he does not find acceptable. (There are all sorts of quibbles that might be raised here. But perhaps they can be put to one side. "Sufficient unto the day . . ." and all that!) Let's also agree that one main source of evidence used in trying to write a grammar for a language is (as you say in your letter) "our intuitive classifications." So the project is to collect intuitions and construct a set of rules or principles which, taken together, entail that a (large or infinite) set of sentences are grammatical (or acceptable), and that another (large or infinite) set of sentences are not. "That," you suggest in your letter, "is how grammar books get written."

We're not sure that that's how *traditional* grammar books get written, since it is far from clear that traditional grammar books were intended to be descriptive rather than prescriptive. (By "traditional grammar books," what we have in mind is the sort of books with titles like "French Grammar for University Students" that we used in university courses aimed at teaching us to speak French.) But that's a quibble. Let's put it to one side. There is a much more important concern about *traditional* grammar books.

As Chomsky argued in some of his earliest works, traditional grammar books are not *fully explicit*. They don't contain enough information to enable a reader to deduce whether a specific sentence is or is not grammatical; to use them to master a language, a reader must often exploit his or her own natural language capacity. We believe that Chomsky is clearly right to claim that traditional grammar books are not fully explicit. His solution was to try to make them fully explicit by writing generative rules and principles which would entail, for every string of words (or sounds), whether or not it was grammatical in the language. So far, we hope, there is nothing here you would disagree with. You would, we expect, agree that traditional grammars were not fully explicit, and you would agree that grammars in the generative grammar tradition aimed to be more explicit, and that (typically) they were.

But now things start to get interesting. For what was discovered as generative linguists tried to write fully explicit sets of rules and principles that would (inter alia) entail which sentences are and are not well formed in a language, is that those rules and principles had to invoke *very* abstract, theoretical constructs. The concepts of noun, verb, clause, etc., familiar from traditional grammar were nowhere nearly adequate (which is not to say they were not useful; they were used in *some* generative grammars). To be fully explicit, generative grammarians found that they had to use concepts like *C-command*, and *X-bar*, and lots of others. There is still lots of dispute among grammarians about which concepts need to be used in writing grammars of natural languages. But just about everyone in that line of work agrees that the job will require *lots* of very technical theoretical concepts which are quite challenging to explain and master. Indeed, one of us (Steve) used to follow that literature, but gave up because he decided that it was just too much work to try to master the increasingly difficult technical notions that were being used in contemporary generative grammars. Moreover, we believe that some of those technical concepts are *so* difficult to master that many perfectly competent speakers of English could not do it if they tried. Steve's grandma was a wise and wonderful lady. But she simply did not have the intellectual wherewithal to understand C-command, any more than she had the intellectual wherewithal to understand some of the more abstract concepts used in quantum mechanics or transfinite recursion theory.

Of course it might turn out that contemporary generative grammarians have just got it wrong. Perhaps there is a way of describing the set of sentences that English speakers judge to be grammatical (or acceptable) that does not require fancy abstract concepts like C-command. But let's assume that they are not radically mistaken in this way. That assumption allows us to ask some questions which we think are crucial in understanding your view: Suppose that GE is a "descriptively adequate" grammar of English. For every sentence, *s*, it entails that *s* is a sentence in English, or that it is not. And these entailments match (near enough) the intuitive judgments that speakers offer about these sentences. Suppose, further, that GE makes use of technical notions like C-command, etc., which many speakers of English could not understand even if they tried to study generative grammar for a few years.

A question:

Q7 Would you say that GE is an implicit theory for speakers like Steve's grandma? Terminology here is awkward and unsettled. If the answer is yes, would you also say that Grandma *holds* GE implicitly? Would you say that Grandma has *tacit knowledge* of GE?—We use these locutions more or less interchangeably, but perhaps you do not.

If you don't think that GE is an implicit theory for Steve's grandma, then we are deeply puzzled. What implicit theory *did* you have in mind when you wrote in FJ[1]: "There is an implicit theory that drives our classifications of sentences in languages we have mastered into the set of the acceptable and the set of the nonacceptable sentences"?

If, on the other hand, you do think that GE is an implicit theory for Grandma, then other puzzles loom. For, as linguists and philosophers of linguistics have been pointing out since the dawn of generative grammar, if there is one "descriptively adequate" grammar of a natural language (as that notion was defined above) then there are many (indeed, probably infinitely many!). A grammar is analogous to an axiomatized theory (with the sentences as the analogues of theorems). And just as there are lots of ways to axiomatize a theory, so too there are lots of sets of generative rules and principles that will "generate" the same set of sentences. Let GE*, GE**, etc. be sets of generative rules and principles, each of which "captures our intuitions" about which sentences are grammatical (or acceptable) in English. (And, suppose also, that GE, GE*, and GE** are roughly equally "simple" and "elegant.")

Q8 Would you say that GE, GE*, and GE** are all implicit theories for speakers like Steve's grandma?

We're not making any bets on how you would answer this. But, as you know, in the Chomskian tradition (which was a major influence on the emergence of contemporary cognitive science), linguists did not rest content with the idea that GE, GE*, and GE** were all implicit theories for Grandma (or, as they preferred to say, they were not all "tacitly known" by Grandma). They insisted that (at most) one of these was the *right* grammar, even though all of them "captured the intuitions." Which one is it? Their answer was that it is the grammar that *is actually represented in the speaker's mind*. There's lots of theoretical baggage in the background here: The mind is (a bit) like a computer. There are rules that really are stored in the mind (just as there are rules that are really stored in a computer). And those rules play a causal role in producing the intuitions. If GE is the right grammar, and GE* etc. are not, that is because (a representation of) GE really is stored in the speakers' minds and really does *play a causal role* in producing intuitions (and comprehension, etc.), while GE*, etc. do not. The passage we just quoted from FJ[1] suggests that you *might* have some sympathy with this idea. For you talk about an "implicit theory that *drives* our classifications" and "drives" certainly invites a causal interpretation. So it is time for another question:

Q9 Do you think that the right implicit theory to attribute to Grandma is the one that is really represented in her mind, and which plays a causal role in generating intuitions?

If the answers to Q8 and Q9 are both *no*, then we're flummoxed. What *do* you mean when you attribute an implicit theory to Grandma? But if the answer to Q8 is *no* and the answer to Q9 is *yes*, still more puzzles await. For if the right implicit theory to attribute to Grandma is the one that is really represented in her mind, then there is good reason to suppose that *merely* probing Grandma's intuitions will not be enough to tell us which one it is. That's useful data, to be sure. But it is not likely to help much in deciding between GE, GE*, GE**, etc. To do that, linguists and cognitive scientists have insisted, we have to look at lots of other sorts of data. Chomsky, famously, thought we have to look at other languages, in an effort to find linguistic universals. (One of us—Steve—used to make his living arguing that this strategy was not likely to succeed, though he has since come to have doubts about his earlier view.) Other linguists and psycholinguists have proposed other strategies, including looking at developmental data and at data from people with impaired language capacities, in an effort to see what the rule system looks like when it is broken or not fully in place. Still others have developed various experimental techniques, using reaction times, Stroop effects, and a whole bunch of other tricks. More recently, some have begun to use brain imaging data in an effort to figure out which rules are actually represented in the mind. No one doubts that it will be hard work to figure out what rules and principles are actually causally responsible for intuitive judgments about grammaticality (etc.). But it seems clear (to us, at least), that once one makes the "realist" move which maintains that the right grammar to attribute to Grandma is the one which is actually represented in her mind, then all sorts of evidence become potentially relevant. If one is a "realist" about grammar, determining the correct grammar for a language (or a speaker) is not something that can be done from the armchair.

Finally, let's return to the distinction between personal and subpersonal levels which we find so puzzling. Suppose, again, that your answer to Q9 is *yes* and further suppose, for sake of argument, that GE is the implicit theory really represented in Grandma, and the theory that plays a causal role in generating intuitions. Our final question is this:

Q10 Would you call GE a personal-level theory, or a subpersonal-level theory?

We suspect that you would have to place GE at the personal level, based on the passage we've already quoted from FJ[1]. There the context indicates that you're talking about personal-level theories, so it seems that what drives our classifications is the implicit personal-level theory. If Chomsky's right, then the theories codified in traditional grammar texts are far too impoverished to drive our classifications, whereas, ex hypothesi, GE does drive our classifications. Thus it seems that, on your account, GE is a personal-level (implicit) theory.

But then we don't know what's going to be left at the subpersonal level. It can't be that subpersonal theories are those which aren't readily accessible to consciousness, since you do seem to allow personal-level theories which aren't readily accessible to consciousness either (i.e., implicit personal theories). Nor can it be that the subpersonal theories are those which refer to properties unfamiliar to Grandma. For C-command structures are unfamiliar to Grandma, and we are assuming that you would nonetheless place GE at the personal level. Even if you don't spot us this assumption, you do allow that personal-level theories might be implicit, and discovering them might be "hard work," which suggests that they might involve unfamiliar properties. So what could be left at the subpersonal level?

We believe that just about all of what we've said about grammar could be said, mutatis mutandis, about folk psychology. But that's best left for another letter. This one has gone on long enough.

All the best,
Kelby and Steve

FJ[2]: Letter from Frank Jackson [2], May 7, 2005

Dear Kelby and Steve,

Your letter of 5 May helps a lot—many thanks. I see that there are a number of key terms that get used differently by different players and that's been the source of some of the disagreement and a fair bit of confusion. This will be a letter mainly about how I'm using the terms. I'll set my comments against some of your questions.

In K&S[2] you say: "As you use the word 'theory,' a theory can be a very small packet of information (or misinformation)." Yes. I was moved many years ago by Quine's decision to use 'object' very inclusively and when it was important to be exclusive, to use 'unified object' or 'object of interest' or. . . . I follow the same policy with 'theory'. You are right that many don't follow this policy and use 'theory' for something that is by definition complex and posits entities that play various roles. But I note that many also say things like "Folk psychology is a complex theory that posits . . ." as if the words 'complex . . . that posits . . .' added something. I agree that it would be good to be explicit about the usage question to avoid needless confusion.

Also in K&S[2] you say: "If, in claiming that FP is not in the head, you were claiming something akin to what many other philosophers have meant when they say that beliefs ain't in the head, then we are no longer puzzled by your claim.—Progress!" Yes I think we have progress but let me sound a word of caution. When people say that beliefs, in the sense of belief contents, ain't in the head, they can have two different things in mind (well, many, but two relevant here). One is that content per se ain't in the head. Belief content is how things are being represented to be and that's not in the head. This is what I had in mind when I mentioned Bob Stalnaker. The other is that belief content need not be shared between doppelgangers in our world (not that there are any of course). I agree with the first claim but deny the second. Many hold that the first entails the second—Bob Stalnaker is an example I think—so the difference matters little for them but it matters for me. How can I deny the clear message of Twin Earth etc.? I'd send you the key papers except that we've enough on hand.

You raise the question in K&S[2]: "Q5 [Are 'explicit' and 'implicit'] the equivalent of '(readily) accessible via introspection' and 'not (readily) accessible via introspection?'" No; but let me answer the next question before I explain what I mean: "Q6 Are we right that you do *not* consider the personal/subpersonal distinction to be the same as the accessible-via-introspection/not accessible-via-introspection distinction?" Yes, but the bald answer may confuse. Let me first say what I mean by the personal/subpersonal distinction. I'll use the sound location example.

When someone hears a sound as coming from location *L* and accepts that things are as their perceptual experience represents them to be, then they believe the sound is at *L*. I say that is representation at the personal level simply to flag that it is the content of what they believe. There's nothing more to the term 'personal' than that. There is a connection with introspection in that people are often rather good at accessing what they believe. But in cases where they do not know what they believe—self-deception, etc.—I'd still say that their belief that *p* was personal level even if they had no idea that they believed that *p*. (I doubt if anything important for our discussion hangs on this last point.)

When someone hears a sound as coming from location *L*, their brains carry the putative information that the sound comes from *L* by virtue of carrying out-of-phase information *PI* indicative of location *L*. I say this is subpersonal information and representation at the subpersonal level because (1) they do not believe that *PI* (nor do they perceptually represent that *PI*), and (2) the information is stored in the brain. The connection with introspection is simply that subjects' failure to access *PI* without fancy brain science is good evidence that they do not believe that *PI*. I suppose that it might be argued that they do believe that *PI* and this is one of the cases where belief content is inaccessible (ditto for perceptual representation), but this would seem a very strained position.

Now for what I mean by the implicit/explicit distinction. It is all to do with the availability to the subject of sentences that capture what they believe—of, that is, sentences that represent as their minds do when they believe that *P*, where *P* is the theory we are talking about. Of course some philosophers have a highly linguistic conception of belief that means that there must always be a suitable sentence available but I belong to the party that thinks that dogs have beliefs and that we have many beliefs that outrun our linguistic capacities—obvious examples being our perceptual beliefs about color shades.

Now for the grammar example. When presented with sentences of English: S1, S2, . . . I am able to classify them as grammatical or the opposite. Let's pretend I am infallible to avoid irrelevant complications. What do I believe when I believe that S17 is grammatical? Three answers might be offered. (1) I believe that S17 is one of the sentences that is tagged 'grammatical' by competent users of English. (2) I believe that S17 plays the "OK" role in its language. (3) I believe that S17 exemplifies the pattern (disjunctive pattern) that unifies the "OK" sentences in English. One way to highlight the difference between (2) and (3) is to note that playing the OK role in its language is shared between grammatical sentences in English and grammatical sentences in Japanese, but the pattern is not.

We know there must be a pattern that unifies the grammatical sentences in English, otherwise we could not acquire the ability to recognize the grammatical sentences in English after a finite number of presentations. Ditto for Russian etc. (Linguists have of course many interesting things to say about the paucity of the evidence base and what that does or does not tell us about the brain, evolution, etc.)

Now consider my belief that S17 is grammatical in sense (3), the sense in which I believe that S17 falls under a certain complex pattern that I can recognize. Do I know what the pattern is? In one sense no. I cannot give an open sentence '*X* is . . .' which represents that *X* has the pattern (or cannot give a set of open sentences, but let's think in terms of a single complex open sentence) that captures its nature. All I can do is produce is '*X* is grammatical'. On the other hand I know case by case what makes a sentence fall under, or fail to fall under, the pattern. I am not in the position of someone who says "I know 'She are happy' is crook but search me how to fix it." I can say case by case for each crook sentence how to fix it (there will be many ways of course) and that case-by-case information is enough to construct in principle the sentential representation. That's what I mean when I say that I have an implicit knowledge of grammar. My situation with respect to grammar is different from the situation of chicken sexers (assuming that the story philosophers tell about chicken sexing is correct).

Now some comments that bear on your good questions.

(1) The case is not one where I believe but am unaware that I believe. I believe that S17 is grammatical (in sense 3), and I believe that I so believe. What I cannot do is produce an open sentence—other than '*X* is grammatical'—that gives the pattern.

(2) What's implicitly known is the pattern, not the open sentence. What's implicitly known is what it is to be grammatical, and that's the pattern and not the sentence. This is why I do not discuss the very interesting question as to whether or not some open sentence that does the job contains terms expressing concepts that I do not possess.

Consider a child who can recognize circles and can do more than merely recognize them. For each plane figure they can say whether or not it is a circle and why—"It's that bump that rules it out." If you draw plane figures on infinitely stretchable rubber, they can distort the rubber to make them circles, etc. What they cannot do is produce '*X* is a circle iff *X* is a plane closed figure with the maximum area to perimeter value', or '*X* is a circle iff *X* is a plane closed figure with such and such a tangent property', or '*X* is a circle iff *X* is a plane closed figure whose perimeter is everywhere equidistant from a single given point', etc. This child will have an implicit grasp of what a circle is by my lights but may well not have the concept of, say, a tangent.

(3) When I say that "[t]here is an implicit theory that drives our classifications of sentences in languages we have mastered into the set of the acceptable and the set of the nonacceptable sentences," I mean that it is the pattern that drives the classifications—not some set of sentences. Of course there will be two patterns—the one in the sentences and the one in the brain. This follows from the fact that the classifications go via the brain. I am talking about the one in the sentences.

There is an interesting question about the pattern in the brain that can be raised in terms of the possibility (the one you mention, and as you say it is more than a possibility) that there are a number of open sentences deploying different concepts that are equally good at picking out the pattern that is being grammatical. I'll raise it in terms of the circle example, as the same applies in that case.

There will be a story about how the child's brain stores the information that some closed figure is a circle, and the different open sentences of the form '*X* is a circle iff . . .' may well differ in how closely they mirror how the brain does the job. Perhaps the brain latches onto the tangent property, or the area property, or the equidistant from a given point property, . . . These are the kind of issues David Marr talks about of course.

(4) One might say that if there are different open sentences that capture the pattern that is being grammatical, there are different patterns in nature that are equally candidates to be the pattern I have an implicit grasp of. But consider the circle again. There are not different shapes for each open sentence. I say the same in the grammar case.

(5) One might point out correctly that often people are wrong about what drives classifications. But I do know what makes me say 'She are happy' is not grammatical. As I say above, the case is not like the chicken sexing case.

Best to you both and hope this helps—pity we can't do this "across a table" but email helps.

Yours,
Frank

K&S[3]: Letter from Kelby Mason and Steve Stich [3], June 18, 2005

Dear Frank,

One of our motives for initiating this correspondence was the suspicion that some of the apparent disagreements between us on issues in the philosophy of mind might be traceable to the fact that we use some crucial terms, like ‘folk psychology’ and ‘tacit (or implicit) theory’, in very different ways. Your last letter has reinforced this suspicion. But it has also raised lots of questions about how, exactly, you do use these terms. We’ll devote this letter to setting out these questions.

The bulk of your last letter was devoted to explaining what you mean by two distinctions: the personal/subpersonal distinction and the implicit/explicit distinction. By and large, we found your account of the first of these distinctions to be unproblematic (with a caveat or two to be noted). By contrast, we found your account of the second distinction to be deeply perplexing. We’ll start with the personal/subpersonal distinction and then go on to the implicit/explicit distinction.

1 Personal/Subpersonal

In FJ[2] you say:

> When someone hears a sound as coming from location *L* and accepts that things are as their perceptual experience represents them to be, then they believe the sound is at *L*. *I say that is representation at the personal level simply to flag that it is the content of what they believe*. There’s nothing more to the term ‘personal’ than that. There is a connection with introspection in that people are often rather good at accessing what they believe. But in cases where they do not know what they believe—self-deception, etc.—I’d still say that their belief that *p* was personal level even if they had no idea that they believed that *p*. (I doubt if anything important for our discussion hangs on this last point.) (FJ[2], emphasis added)

So, if we understand you correctly, a *personal-level X* (for a given person, *P*) is simply an *X* that is the content of (one or more of) *P*’s beliefs. One question that arises here is: What kinds of things can be personal-level *X*s? Since a personal level *X* must be a content of a belief, we assume that the answer is that all personal-level Xs are *representations* or *contents*—when these are understood, as explained in FJ[1], as things that are not in the head. Since you identify a theory with its contents, you can also talk about a personal-level theory. (Indeed, given your broad use of ‘theory’ every personal-level *X* is a personal-level theory.)

Before turning to the more problematic implicit/explicit distinction, we want to raise three concerns about your account of the personal/subpersonal distinction. The first of them will probably have no bearing at all on the sorts of issues we’ll be discussing in this correspondence; we raise it because it might be relevant to understanding some of your views on other matters. If the notion of the personal level is tied to *belief* in the way we’ve stated above, then there can’t be personal-level representations that are not propositional (or, if you prefer a less jargon-laden way of making the point, there can't be personal-level representations that are not possible contents of beliefs). So, for example, since one can't *believe* a concept, there are no personal-level concepts. And since one can't *believe* the content of a taste sensation, the contents of taste sensations can't be personal-level things either.

Is this really the way you want to use the notion of the personal level? We rather suspect that it isn't, and that your explanation of the personal level, which appeals to the contents of *what is believed*, was intended only for the special case of propositional representations or contents. That's fair enough, though it does leave you with the job of giving an account of the distinction between personal-level and subpersonal-level representations when those representations are not propositional.

A second concern, closely related to the previous one, is that the interpretation we've proposed of the passage from FJ[2] is almost certainly too literal minded in restricting itself to *belief*, since your personal/subpersonal distinction would presumably hold for attitudes like desires as well. For instance, suppose Waugh wants to throw the ball at the stumps, in order to run out Muralitharan. One representation in Waugh's mind is "I [Waugh] throw the ball at the stumps." At the same time, his mind will also contain some representation with more motoric detail, along the lines of "I throw the ball with a velocity of at least *x*, so that it traces a trajectory of *y* . . ." We suspect that you would call the first representation personal and the second subpersonal. So a more charitable interpretation of the passage we quoted would be: a personal-level *X* is the content of one or more of *P*'s beliefs or desires (or perhaps his propositional attitudes in general?). We raise this concern just to make sure we've understood your distinction between the personal and subpersonal levels; in this letter we'll continue to focus on belief as the attitude we're all interested in.

Our third concern is that tying the notion of the personal level to the notion of belief could lead to some question-begging mischief down the road. Why? Well, one of the reasons we are all interested in getting clear about the notions of *implicit theory* and *folk psychology* is that these notions play a central role in the debates over eliminativism. And eliminativists, of course, maintain that there are no beliefs. If they're right, then on your account of the personal level, there are no personal-level representations either. So if one is going to make claims about personal-level representations in laying the groundwork for the eliminativism debate—in saying what folk psychology is, perhaps—then one has to be very careful not to beg the question. The devil is in the details, of course. Whether a question is indeed being begged will turn on the details of the argument at hand. All we want to do here is to raise a caution flag.

2 Implicit/Explicit

We turn now to your account of the implicit/explicit distinction, which, as we mentioned earlier, we find very perplexing. To explain why we're perplexed, we'll raise four (clusters of) problems, going from less to more serious.

Q11 To what sorts of things does the distinction apply?

As in the case of the personal/subpersonal distinction, we are not entirely clear about the sorts of things to which you would apply the terms 'implicit' and 'explicit'. The clear case, of course, is *theories*. You say: "there is an *implicit theory* that drives our classifications of sentences in languages we have mastered into the set of the acceptable and the set of the nonacceptable sentences" (FJ[2], emphasis added).

And elsewhere you make it clear that you think folk psychology is also an implicit *theory*. So theories, which, for you, are propositional (?) representations or contents, or clusters of such representations, are one kind of thing that can be implicit or explicit. But in FJ[2] you also talk, on several occasions, of implicit knowledge of a *pattern*. ("What's implicitly known is the pattern . . ."; "there are different patterns in nature that are equally candidates to be the pattern I have an implicit grasp of.") Later in this section, we'll raise some concerns about your notion of a pattern. But whatever a pattern is, it seems unlikely that patterns are *theories*. (One can't believe a pattern; it seems odd to say that a pattern is true or false.) So there are at least two kinds of things to which you would apply the terms 'implicit' and 'explicit'. Are these the only two? Or are there more?

Q12 What is "availability"?

Here is the passage from FJ[2] in which you give your most explicit account of the implicit/explicit distinction:

> Now for what I mean by the implicit/explicit distinction. It is all to do with the availability to the subject of sentences that capture what they believe—of, that is, sentences that represent as their minds do when they believe that *P*, where *P* is the theory we are talking about. Of course some philosophers have a highly linguistic conception of belief that means that there must always be a suitable sentence available but I belong to the party that thinks that dogs have beliefs and that we have many beliefs that outrun our linguistic capacities—obvious examples being our perceptual beliefs about color shades.

We'll shortly raise some questions about just what *kind* of sentence has to be available for a theory (or pattern) to be explicit. But first we want to ask what you mean by 'availability'. There are three ways to unpack the idea that there are sentences *available* to a subject which capture what he believes. In the strong sense, a subject has such sentences available when he can *produce them* if asked (with, perhaps, some idealization about full rational reflection and the like). In the moderate sense, the subject need only be able to recognize them as sentences which do indeed capture what he believes. In the weak sense, all that's needed is that there be sentences in the subject's language which will do the job, even if he couldn't produce them when asked, or recognize them as being the right sentences.

We assume that when you talk of sentences being available, you mean available in the *strong* sense. The weak sense is obviously too weak, since it would make too many theories explicit. The moderate sense also seems too weak. To see why, take your example of the child who can tell circles from noncircles but who has no beliefs of the form '*X* is a circle iff *X* is a plane closed figure with the maximum area to perimeter value'. Let's assume that the child—call him Meno—has had some mathematical education, so he understands all the key terms in this sentence—'plane closed figure', 'maximum', 'area', 'perimeter'. Suppose now someone tells Meno the sentence just quoted, viz. "*X* is a circle iff . . ." and, after quick reflection, Meno agrees. In the moderate sense of 'available', this sentence was available to Meno and thus Meno had an explicit theory of circles. Since you *don't* think someone like Meno (before his enlightenment) has an explicit theory of circles, we infer that the moderate sense of 'available' is also too weak.

Though we think that the strong sense of 'available' is what you intended, your examples of dogs and perceptual beliefs convinced us that perhaps we had best ask whether this is right. For in both these cases the required sentences are not available in the strong (or moderate) sense because they are not available in the weak sense. So, before going on, we want to be sure we understand you. When you talk of appropriate sentences being available, do you, as we assume, interpret 'available' in the strong sense, viz. there not only have to be suitable sentences in the subject's language, but the subject has to be able to produce them when asked to produce a sentence which captures what he believes?

Q13 What kind of sentences have to be available?

OK. Now that those preliminary concerns are on the table, let's turn to what we find most problematic about your account of the implicit/explicit distinction. To begin, let's quote, again, the crucial passage from FJ[2]: "Now for what I mean by the implicit/explicit distinction. It is all to do with the availability to the subject of sentences that capture what they believe—of, that is, sentences that represent as their minds do when they believe that *P*, where *P* is the theory we are talking about." So if someone can produce sentences which "capture" what he believes, then his theory is explicit, while if he cannot, then his theory is implicit. And to "capture" what he believes, the sentences have to "represent as their minds do when they believe that *P*, where *P* is the theory we are talking about." Right after this passage, you apply the idea to the example of grammar. And it is there that we get well and truly lost. Here is part of what you say: "Now consider my belief that S17 is grammatical in sense (3), the sense in which I believe that S17 falls under a certain complex pattern I can recognize. Do I know what the pattern is? In one sense no. I cannot give an open sentence '*X* is . . .' which represents that *X* has the pattern (or cannot give a set of open sentences, but let's think in terms of a single complex open sentence) that captures its nature" (FJ[2]).

Now one might think that it is *trivial* to give the sort of open sentence you require, since surely '*X* is *grammatical*' is an open sentence which represents that "*X* has the pattern . . . that captures its nature." But clearly *that* is not what is required, since your very next sentence is: "All I can do is produce is '*X* is grammatical.'" But if '*X* is grammatical' won't do, why not? Surely it does "represent that *X* has the pattern that captures its nature." So we infer that there must be some additional constraints on the open sentence that must be available. But what could those constraints be? Here it would be easy enough to generate a long list of uncharitable proposals, and to show that they won't work. Perhaps the most obvious candidate is the constraint that the sentence can't use the word 'grammatical'. OK. Then you can say '*X* is a sentence and *X* is not ungrammatical'. Or '*X* is a well-formed sentence in my idiolect'. Or '*X* is a sentence in the dialect of English that I speak'. Or '*X* is a sentence generated by the internally represented set of linguistic rules that plays a central role in my language processing'. Or . . . But there is really no point in offering uncharitable proposals and showing that they won't work. Our goal isn't to argue against your view but to *understand* it. And here, we must confess, we are just flummoxed. So let us try to pose the question that puzzles us simply and clearly as possible:

Q14 What *are* the constraints on the "belief capturing" open sentence that must be available to a person if his theory is to count as explicit rather than implicit?

Before leaving this section, it might be useful to offer a very rough sketch of the account of explicitness that, we believe, plays an important role in linguistics and psycholinguistics. As we noted in K&S[2], early on in the history of generative grammar, Chomsky argued that traditional grammar books were not fully explicit; to determine what they claim about a specific sentence (or sequence of words) in the target language, one often has to make sophisticated use of one's own linguistic capacity. To make grammars fully explicit, Chomsky proposed that we try to discover a set of generative rules and principles which would entail, for every string of words (or phonemes), whether or not it was well formed in the language.[5] The notion of entailment invoked here was closely related to the notion of entailment assumed in math and computer science. Roughly speaking, for every grammatical sentence in the language, there has to be a formal proof from the rules and principles to the claim that '*S* is grammatical'. And, even more roughly speaking, the proof is formal in the sense that its correctness can be checked by a computer. Back in the old days when Steve was young, a grammar that was fully explicit and that correctly specified the set of sentences in the target language was said to be "descriptively adequate." But, as we noted in K&S[2], for a variety of reasons people in the Chomskian tradition were not satisfied with the goal of producing grammars that were merely descriptively adequate. Some of them, including Chomsky, insisted that we should aim at discovering the grammar that is actually represented in the minds of speakers of the language and which plays a causal role in producing linguistic intuitions, and in more important jobs like comprehension and speech production.

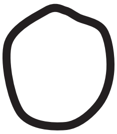
Now in the quote from FJ[2], you say: "It is all to do with the availability to the subject of sentences that capture what they believe—of, that is, sentences *that represent as their minds do when they believe that P*, where *P* is the theory we are talking about." And taken out of context, the passage we have emphasized might be read as claiming that what must be available is what the Chomskians want—viz. a statement of the rules that are actually represented in the speaker's head. But in light of everything else you say, we take this interpretation to be unlikely. Another interpretation of your view on what must be available for a theory to be explicit, is that what is required is something like what the Chomskians call a descriptively adequate grammar. That may be closer, but we doubt you'll embrace that proposal either. So, rather than speculating further, let us bounce it back to you. How should your account of the explicit/implicit distinction be interpreted? What is the answer to question Q14?

Q15 What are "patterns"?

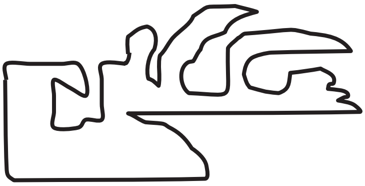
So far, we have said very little about "patterns," though they loom large in your discussion of the implicit/explicit distinction. Here, again, we find what you say quite puzzling. We'll begin by noting what *might* just be a slip on your part, though it might reflect some much deeper point that we haven't understood. In discussing implicit knowledge of grammar, in FJ[2], you say: “I know case by case what makes a sentence fall under, or fail to fall under, the pattern. I am not in the position of someone who says ‘I know “She are happy” is crook but search me how to fix it.’ I can say case by case for each crook sentence how to fix it (there will be many ways of course) and that case-by-case information is enough to construct in principle the sentential representation.”

Now it is surely true that, presented with some cases of “crook” sentences, you know how to fix them. But, if “crook” means *ungrammatical*, then we think it is simply false that, for *each* crook sentence, you know how to fix it. Every sentence in Hungarian is an ungrammatical sentence in English. But presented with a sentence in Hungarian and asked to “fix it,” we would be stymied. We would have no idea what we were supposed to do and, we assume, you wouldn’t either. Perhaps you want to restrict your claim to “crook sentences” that are made up of English words. But even here, we think your claim is false. It is easy enough to “fix” your example, ‘She are happy’. But how would you “fix” something like ‘The the the in loud a a’? Like you, we judge this sequence of words to be ungrammatical. But asked to fix it, we’d be no better off than we would be in the case of the Hungarian sentence.

A bit later, you make some related remarks about “a child who can recognize circles and can do more than merely recognize them. For *any plane figure* they can say whether or not it is a circle and why—‘It’s that bump that rules it out’ ” (FJ[2], emphasis added). Here again, we’re puzzled. There are no doubt lots of kids who can do just what you say for figures like this one:



But what is the kid supposed to say when asked why the following figure is not a circle?



If we were asked this question, we would not have a clue about how to answer (unless: "Because it isn't round!" counts as an answer).

Is any of this of any importance? Well, frankly we're not sure. What is clear is that the examples of the "crook sentence" and the figure that is not a circle loom large in your discussion of patterns. Moreover, patterns play a central role in your account of what is implicitly known in the case of grammar: "(2) What's implicitly known is the pattern, not the open sentence. What's implicitly known is what it is to be grammatical, and that's the pattern and not the sentence" (FJ[2]). And you seem to think that the grammar case and the circle case are deeply analogous.

We find all of this deeply puzzling. To explain why, let's focus on the case of grammar. In one example in FJ[2] you say: "I believe that S17 exemplifies the pattern (disjunctive pattern) that unifies the OK sentences in English." And a bit later, you elaborate on this as follows: "We know there must be a pattern that unifies the grammatical sentences in English, otherwise we could not acquire the ability to recognize the grammatical sentences in English after a finite number of presentations. Ditto for Russian etc."

But what could it mean to say that there is a *pattern* that unifies all the grammatical sentence of English? Well, one thing that is surely true is that there is a *set* of sentences (or phoneme sequences) which are grammatical in English. Another undisputed truth is that there infinitely many sentences in that set. And it is widely held that considerations of learnability like those you cite indicate that the set of grammatical sentences must be recursive (or, a bit more modestly, that it must be possible to characterize them with a finite set of rules or principles). Finally, as we noted in our previous letter (K&S[2]) if there is one finite set of rules that generates a given infinite set, there are indefinitely many. As far as we can see, these are all the facts there are in this vicinity (as Dave Chalmers likes to say). So when you talk about a pattern that "unifies" the grammatical sentences in English, which of these facts do you have in mind? Is it just a way of saying that the sentences of English are members of an infinite recursive set (or an infinite set that can be characterized by a finite set of rules)? If that's *not* what you mean, then we need some help, since we can't think of anything else that you might mean. But if that *is* what you mean, then problems loom elsewhere.

For consider (2) above. You say that what is implicitly known is the pattern. On the reading we're proposing, what that means is (something like) what is implicitly known is an infinite set that can be characterized by a finite set of rules. So far, so good. Although we should note in passing that 'implicit knowledge' in this context strikes us as a bit odd, since to say that a person implicitly knows the set (= the pattern) in this context is simply to say that she can (under idealized circumstances) correctly judge that each sentence in the set (and only these) is grammatical. And that is simply an observation about how the person *behaves* (under idealized conditions). One might have thought that positing implicit knowledge was a way of *explaining* a range of behaviors; but on this reading, talk about implicit knowledge does no serious explanatory work at all. It is just a way of *describing* the behavior that the subject would exhibit.

That this is more than just a terminological curiosity emerges in the following quote: "(3) When I say that '[t]here is an implicit theory that drives our classifications of sentences in languages we have mastered into the set of the acceptable and the set of the non-acceptable sentences,' I mean that it is the pattern that drives the classifications—not any set of sentences" (FJ[2]). (The context suggests that in the last bit, "not any set of sentences," you're talking about the "open sentences" discussed earlier, which, are not "available" to subjects in cases of implicit knowledge. Though nothing we say will turn on this reading.)

Now, as we noted in K&S[2], 'drives' invites a *causal* interpretation. And here we are completely stumped. How could it be true that an infinite recursive set (or an infinite set that can be characterized by a finite set of rules) *drives* (i.e. causes) our classification of sentences? What does the driving (at least if anything like the Chomskian story is correct) is the specific set of recursive rules that is represented in the speaker's brain. The sentence that follows (3) leaves us even more confused: "There is an interesting question about *the pattern in the brain* that can be raised in terms of the possibility (the one you mention, and as you say it is more than a possibility) that there are a number of open sentences deploying different concepts that are equally good at picking out the pattern that is being grammatical" (FJ[2], emphasis added). Here again, the reading of your "pattern"-talk that we proposed above makes no sense. If a pattern is an infinite recursive set, then we're reasonably confident that there are no patterns in our brains since our brains are distressingly finite.

3 Conclusion

When we started to write this letter, we planned to end with a sketch of the account of "tacit" theories that we favor,[6] and compare it to your account of "implicit" theories. But for two reasons, we now think that is best saved for another occasion. First, we don't really understand your account well enough to do a *compare and contrast*. Second, this letter is already very long. So we'll close with the hope that your response will

dissolve some of the puzzlement we've expressed about your account of the explicit/implicit distinction.
All the best,
Kelby and Steve

FJ[3]: Letter from Frank Jackson [3], June 21, 2005

Dear Kelby and Steve,
I'm replying quickly to your letter of 18 June in the hope of heading off some misunderstandings.

1 Personal versus Subpersonal

I illustrated this distinction with the case of belief because that is what we've been discussing. I don't think that personal-level states are one and all beliefs. I think being a belief is sufficient for being a personal-level representational state but not necessary—and I take it this is standard doctrine. I think, for example, that perception is a personal-level representational state, and one can perceptually represent that *p* without believing that *p*.

It may also help if I comment briefly on "your three concerns about your [my] account of the personal/subpersonal distinction": "If the notion of the personal level is tied to belief in the way we've stated above, then there can't be personal-level representations that are not propositional (or, if you prefer a less jargon-laden way of making the point, there can't be personal-level representations that are not possible contents of beliefs). So, for example, since one can't believe a concept, there are no personal-level concepts. And since one can't believe the content of a taste sensation, the contents of taste sensations can't be personal-level things either" (K&S[3]).

I distinguish representational states from representations. I think all representational states are propositional (although not because I believe all representational states are beliefs). I don't think all representations are propositional. I doubt if there's any substantial difference between us here. It is a question of terminology. By the way I think you can believe the content of a taste sensation. When something tastes sweet to me I may well believe that it is sweet: "your personal/subpersonal distinction would presumably hold for attitudes like desires as well" (ibid.).Yes. Again: "Our third concern is that tying the notion of the personal level to the notion of belief could lead to some question-begging mischief down the road" (ibid.). I agree. The personal versus subpersonal distinction would apply even if certain eliminativist views turned out to be correct.

2 Implicit/Explicit

(i) You ask in your previous letter: “To what sorts of things does the distinction apply?” and worry about my talk “of implicit knowledge of a *pattern* . . . [and say] . . . it seems unlikely that patterns are theories. (One can’t believe a pattern; it seems odd to say that a pattern is true or false.) So there are at least two kinds of things to which you would apply the terms ‘implicit’ and ‘explicit’. Are these the only two? Or are there more?” (ibid.).

I think you’ve been foxed by my phrasing. The distinction applies to theories (in my inclusive use of that term). Knowledge of a pattern is knowledge about where and when the pattern is exemplified and this is a theory and can be true or false (and will be true if it is indeed a case of knowledge). Think of belief as to where the party is. You might say that one cannot believe where a party is, and that it seems odd to say that where a party is is true or false. But of course belief as to where a party is is any belief of the form ‘the party is at *x*’ and will be true (false) just if the party is at *x* (is not at *x*). Likewise, when I talk of knowledge about a pattern, I mean knowledge to the effect that *x* falls under the pattern, and it will be true or false that *x* falls under the pattern in question.

(ii) You ask “What is ‘availability’?” and produce three possibilities. Roughly, I mean the sense you dub the strong sense (the quibbles aren’t worth the space).

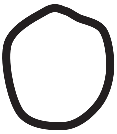
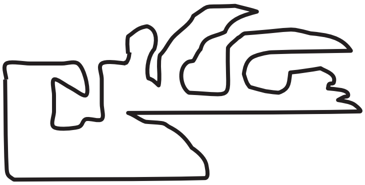
(iii) You ask “*what kind of sentences have to be available*” and make the entirely correct point that we do have in the case of grammar a sentence available that represents that *S* is grammatical, namely, the sentence ‘*S* is grammatical’. How then do I count our knowledge of grammar as a good example of an implicit theory? The answer is that we cannot produce a sentence that elucidates the pattern. We know that being grammatical is not sui generis; we know there’s a structure to being grammatical but we cannot give it in words.

Now it is my turn to be baffled. I cannot see why you find all this so hard. There is a structure to being a wff in logic. Some students who know that being a wff is a structured property can give the structure in words (there’s more than one way to do this but that’s fine). Some cannot but they can reliably recognize wffs and they know case by case why a non-wff is a non-wff. We need some way of describing the knowledge of the second group. I say they have implicit knowledge. Maybe there are better terms; maybe there are better ways of making the distinction, but I cannot see why you resist the point that there is an interesting phenomenon here that calls for a name.

Maybe the answer lies in what you say about my talk of patterns but here I am equally baffled by your remarks. Anyhow, let me go through the cases that puzzle you in the section of your previous letter after Q15, "What are 'patterns'?": "Now it is surely true that, presented with some cases of 'crook' sentences, you know how to fix them. But, if 'crook' means ungrammatical, then we think it is simply false that, for each crook sentence, you know how to fix it. Every sentence in Hungarian is an ungrammatical sentence in English. But presented with a sentence in Hungarian and asked to 'fix it,' we would be stymied." I don't think that this last claim is true. Here's one way to fix it. Put quote marks around it and add 'is not a sentence of English' to the result. That will do the trick. What is true is not that we don't know how to fix it but rather that there is no salient candidate to be *the* way which deviates the least from what we started with.

You go on to say: "Perhaps you want to restrict your claim to 'crook sentences' that are made up of English words. But even here, we think your claim is false. It is easy enough to 'fix' your example, 'She are happy.' But how would you 'fix' something like 'The the the in loud a a'?" (ibid.). Many answers: put it in quotes and append 'is not well formed in English'; replace it by 'The sound is loud'; etc.

You discuss (following Q15 in K&S[3]) my claim saying:

> For any plane figure they [certain children] can say whether or not it is a circle and why—"It's that bump that rules it out." Here again, we're puzzled. There are no doubt lots of kids who can do just what you say for figures like this one:
>
> 
>
> But what is the kid supposed to say when asked why the following figure is not a circle?
>
> 

If we were asked this question, we would not have a clue about how to answer (unless: "Because it isn't round!" counts as an answer).

But it is easy for the child to say why the figure is not a circle. What would puzzle them is a request to pick out the *one* reason—there are lots and lots of reasons. The bent bit at the bottom right, the bent bit at the top left, the flat section below the wave-like bit, etc., etc.

Why is it that we seem to so often to puzzle each other? Maybe the nub of it is in your question "But what could it mean to say that there is a pattern that unifies all the grammatical sentence of English?" (ibid.).

You go on to discuss this as if it were a question about language. For example, you discuss the idea that it means "the set of grammatical sentences must be recursive (or, a bit more modestly, that it must be possible to characterize them with a finite set of rules or principles)" (ibid.). For me it isn't. If there's a pattern, it will in principle be possible to capture it in words and there will be many ways to do this. But the existence of a pattern is a fact about the world—in particular about a certain set of sentences—and not about words. Square things are alike; we can capture the fact in words; but the similarity is not a fact about words. The difference between square things and grammatical sentences is the extent to which the similarity is disjunctive but the sense in which both are cases of things falling under a pattern is the same.

This means I find your puzzlement about my talk of driving itself puzzling. If there's a pattern, why should we not be able to recognize it, and why shouldn't the ability to recognize it do some driving—that is, some causing of sentences like 'The sentence before me is grammatical'?

Or take the following passage from your letter (ibid.): "Finally, as we noted in our previous letter (K&S[2]) if there is one finite set of rules that generates a given infinite set, there are indefinitely many. As far as we can see, these are all the facts there are in this vicinity (as Dave Chalmers likes to say). So when you talk about a pattern that 'unifies' the grammatical sentences in English, which of these facts do you have in mind?" For me there is a glaring omission in the list of facts in this vicinity—you haven't mentioned the pattern in nature. You've left out the main player, as I see things.

All the best,
Frank

K&S[4]: Letter from Kelby Mason and Steve Stich [4], July 11, 2005

Dear Frank,

Many thanks for your letter of 21 June. It's been rough going. But we think we have finally begun to understand how you use the implicit/explicit distinction, though there are still a number of issues on which we need lots of help. Like you, we continue to be surprised at how hard it is for philosophers who work in, near enough, the same philosophical tradition to understand each other. Sometimes we worry that the problem is idiosyncratic to us, and that we're just being dense. If that's the case then this exchange will be of very limited interest. But if, as we suspect, lots of others have seriously misunderstood your claims about folk psychology and implicit theories, then the time invested on this project will be well spent. In this letter, we'll start with the bits we think we understand, then go on to the issues on which we still need help. In the last section, we'll (finally) get back to folk psychology.

1 The Parts We Think We Understand

It was the following rather exasperated paragraph in your letter of 21 June (FJ[3]) that finally made your use of 'implicit knowledge' start to come into focus. (The numbering and emphasis is ours, of course.)

> [A] Now it is my turn to be baffled. I cannot see why you find all this so hard. There is a structure to being a wff in logic. Some students who know that being a wff is a structured property can give the structure in words (there's more than one way to do this but that's fine). *(iii) Some cannot but (i) they can reliably recognize wffs and (ii) they know case by case why a non-wff is a non-wff.* We need some way of describing the knowledge of the second group. I say they have implicit knowledge. Maybe there are better terms; maybe there are better ways of making the distinction, but I cannot see why you resist the point that there is an interesting phenomenon here that calls for a name.

You are certainly right that the sentence we've emphasized characterizes an interesting phenomenon, that it calls for a name, and calling it 'implicit knowledge' is as good a label as any. But since we've misunderstood each other so often in the past, it's important to go very slowly here and be very clear on what the label is a label for. As we read you, it is a label for an *ability* or *capacity* that some people have. And, as we've indicated with our numbering, the italicized sentence suggests that the ability has three components.

(i) The ability to *recognize* wffs—which, we assume, is the ability to tell, for any symbol sequence whether or not it is a wff. (Obviously, this is an idealization, since no one is very good at recognizing wffs that are, say, over ten million symbols long. And there are issues about how the idealization should be unpacked. But we don't think that's worth pursuing here, since there are similar issues about how to characterize just about any open-ended capacity.)

(ii) The ability to *say*, for each symbol sequence that is not a wff, why it is not a wff. (You talk about *knowing* rather than the ability to say, and these are surely different, though for the purposes at hand, we don't think the difference will be important.)
(iii) Lack of the ability to "give the structure in words."

If *S* has these three abilities (or, if you prefer, two abilities and an inability) you'd say: '*S* has implicit knowledge of wffs', or perhaps you would prefer locutions like: '*S* implicitly knows what wffs are' or '*S* implicitly knows what it is to be a wff'. We don't think anything hangs on which of these locutions you'd use. But if you think there is some principled reason to prefer one of them, or some other similar locution, please let us know.

Though the account we've just given was couched in terms of the wff example, it is easy to see how to generalize to many other cases. The most obvious is the one we've discussed in previous letters, viz. grammaticality. If *S* (i) has the ability to recognize grammatical sentences in English, (ii) has the ability to say, for each nonsentence, why it is not a sentence, and (iii) lacks the ability to say what it is to be grammatical, then *S* has implicit knowledge of grammaticality in English (or implicitly knows what it is for a sentence to be grammatical in English). The generalization to another one of your examples, viz. circle, is straightforward as well. How to generalize to other important cases, including folk psychology, is less clear. But that's a topic for Part 2.

Have we got it right so far? Or close to right? (We certainly hope the answer is *yes*. It would be a real bummer to have to start all over again!)

If we have got it right, then we can set to rest one concern that loomed large in previous letters. We have repeatedly protested that we did not understand your claim that an implicit theory "drives our classifications." But as we now understand, for you, a crucial part of having implicit knowledge is having an *ability to recognize*. And, as you put it in FJ[3]: "If there's a pattern, why should we not be able to recognize it, and why shouldn't *the ability to recognize it* do some driving—that is, some causing of sentences like 'The sentence before me is grammatical'?" (Emphasis is ours.)

Later on, we'll have another go at the pattern-talk, which we take to be deeply problematic. But that's not crucial here. What is crucial is that when you talk about implicit knowledge driving judgments or utterances, all you are claiming is that an ability to recognize grammatical sentences, for example, plays a causal role in producing the judgment that a specific sentence is grammatical. Of course it does. While this isn't a very deep or particularly illuminating explanation of our grammaticality judgments, you never said it was. But does it make sense to say that the ability to recognize grammatical sentences *drives* (= plays a central role in causing) grammaticality judgments? Of course it does. Issue resolved.

2 The Parts on Which We Still Need Help

In the previous section we suggested that the interesting phenomenon that you call 'implicit knowledge of *X*s' can be understood as the conjunction of two abilities and an inability. We've got no problem with the first of these—the ability to tell, for any object, whether or not it is an *X*. (As we noted, there are idealization issues, but we're not going to worry about those.) We do, however, have problems with both (ii) and (iii). We've tried to articulate the problems in previous letters, where we had a much less clear view of the role they play in your account of implicit knowledge; perhaps we can do a better job here.

2.1 Let's start with (ii): knowing (or being able to say), for each non-*X*, why it isn't an *X*. One thing that your most recent letter made clear is that for some reason this condition is *important* to you. In our last letter, we said we weren't sure how important it was to get clear on this stuff about knowing (or saying) why a non-*X* isn't an *X* (K&S[3]). But the issue has come up in the last two of your letters, and you spend quite a lot of time on it in FJ[3]. So clearly you do take it to be important. Unfortunately, we are still deeply puzzled about what this condition on implicit knowledge requires, and about why you think it is important to include it in your account of implicit knowledge.

To try to explain our puzzlement, let's go back to the first time this condition (or something in the vicinity) made its appearance in our correspondence. Discussing the example of grammar, in FJ[2], you wrote: "[B] I know case by case what makes a sentence fall under, or fail to fall under, the pattern. I am not in the position of someone who says 'I know "She are happy" is crook but search me how to fix it.' I can say case by case for each crook sentence how to fix it (there will be many ways of course) and that case-by-case information is enough to construct in principle the sentential representation."

Let's assume, for the time being, that this earlier passage is spelling out (ii) as the ability to "fix" non-*X*s. In K&S[3], we protested that while it is reasonably clear what would count as fixing in the case of near misses—almost-*X*s like 'the cat sitted on the mat'—there are plenty of ungrammatical sentences where we have no idea what fixing them would be. The examples we offered were a sentence in Hungarian and the word sequence 'The the

the in loud a a', neither of which is grammatical in English. Asked to "fix" them, we insisted, we wouldn't know what to say. In your most recent letter (FJ[3]) you respond to both cases. Regarding the Hungarian sentence, you write: "Here's one way to fix it. Put quote marks around it and add 'is not a sentence of English' to the result. That will do the trick." And for "The the the in loud a a" you suggest: "Many answers: put it in quotes and append 'is not well formed in English'; replace it by 'The sound is loud'; etc."

Now you are certainly right that each of these strategies will produce a grammatical English sentence. What we find puzzling is the claim that these strategies will "fix" the ungrammatical sentence. You are, of course, free to use the term 'fix' however you see fit. But if you mean these examples seriously, then it seems that, as you use the term 'fix', there are just about no constraints on what counts as fixing. Is it enough to produce a grammatical sentence which either uses or mentions one or more of the words in the original ungrammatical sentence? That seems to be the *only* obvious feature that your three proposed fixes have in common. If that *is* enough to count as a "fix," then the requirement that one must be able, on a case by case basis, to "fix" every ungrammatical sentence turns out to be absurdly weak,[7] and it is hard to see why you want to impose it at all or what work it can do for you. To make the point in another way, consider the last bit of quote [B] above. If fixing is interpreted in the extremely weak way that your examples suggest, why would you think that the information obtained from fixes (in that weak sense) "is enough to construct in principle the sentential representation"? Frankly, this claim just makes our heads spin.

Now in quote [A], which may be a more careful statement of your view, you do not actually impose the able-to-fix-it requirement. So perhaps the able-to-fix-it requirement of [B] isn't really a part of condition (ii). Rather, what you strictly require in [A] is (ii), knowledge (or the ability to say), for each non-*X*, why it isn't an *X*. But here we're back to our old problem. If we interpret this requirement in what we take to be the natural way, then it is absurdly *strong*. If someone gave us a Hungarian sentence and asked us to say why it is not a grammatical sentence in English, we would have no idea what to say (other than "because it isn't English"). Similarly, if someone gave us the sentence 'The the the in loud a a' and asked us to say why it is not a grammatical sentence in English, we would not know what to say. So, interpreting your requirement in what we take to be the most obvious way, it would follow that Kelby and Steve do not have implicit knowledge of grammaticality in English. And this surely is not what you want to say.

There are, of course, lots of other ways in which requirement (ii) could be interpreted. We're not at all sure that you have a clear interpretation in mind. But if you do have something clear in mind, it will have to avoid the twin pitfalls we've noted. It can't be so weak (Give me a good sentence that uses or mentions one of the words in the bad one) that it excludes almost nothing, nor can it be so strong that it excludes folks like us from having implicit knowledge of English grammar.

Let us close this section with a pair of questions (well, OK, Q17 is actually a whole bunch of related questions):

Q16 How, exactly, do you propose to interpret requirement (ii)?

Q17 Why do you *want* to impose a requirement like this in your account of implicit knowledge? What work is it doing for you? Another way of putting this question is: Why don't you just drop requirement (ii) and say that a person has implicit knowledge of *X* when he has the ability to reliably recognize *X*s—as this idea is unpacked in (i)? (We're ignoring (iii) for a moment; we'll get to that next.) We suspect this is the crucial question, since what really puzzles us is why you think you need a requirement like (ii) in addition to (i).

2.2 Let's turn, now, to requirement (iii). Here, in contrast with (ii), we think we understand why you *want* to impose some condition like this, since what this requirement is trying to ensure is that *S* does not have *explicit* knowledge. If *S* has the ability to recognize *X*s and can *also* give a suitably detailed account of what *X*s are, then you want to say *S* has explicit knowledge of *X*s, not implicit knowledge.

But while the motivation is clear, the requirement, unfortunately, is not. We went on at some length about this in K&S[3]. What that discussion led up to was Q14, about the constraints on "belief capturing" open sentences. In your reply (FJ[3]) you never answered this question. The closest you come is to talk about a sentence that "gives the pattern." But for reasons we'll discuss in the section on patterns, we don't think this even begins to count as an answer to the question.

2.2.1 A modest proposal Up until now, in this correspondence, we've mostly been asking questions and trying to understand your view. But we now think we understand your view and its shortcomings well enough to propose an alternative. As we interpret your view, particularly the motivation for (iii), you want implicit knowledge and explicit knowledge to exclude one another. A person can have one or the other sort of knowledge of a theory or pattern or property, but not both. But in order to draw the distinction, you need some clear and well-motivated answer to Q14. And, to be frank, we don't believe that an answer will be forthcoming.

What we propose is a rather different way of thinking about the interesting (and not so interesting) phenomena in this vicinity. There are at least three of them.

(1) The first (basically (i)) is the ability to reliably recognize whether or not something is an *X*—where *X* can be *grammatical sentence in English*, *wff*, *circle*, and lots of other things.

(2) Second, someone who can recognize *X*s can also know how he does it. In some cases (grammar, perhaps) this will involve knowing that his mind–brain uses a *specific* set of rules or principles to compute the answer to the question 'Is this an *X*?'. One of the goals of much cognitive science is to produce knowledge of this sort, both about one's own recognition abilities and about other people's. And while there has been impressive progress in a few areas, there is still much work to do. There is very little of *this* sort of knowledge in the world at the moment. Perhaps none at all.

It is important to note that (1) and (2) are *not* exclusive. A person can have both (1) and (2). Moreover, even if someone has the sort of knowledge required for (2), she often will not use it to recognize *X*s. If an English-speaking grammarian or psycholinguist ever does figure out how we distinguish sentences that are grammatical in English from sentences that aren't, she will rarely if ever use this knowledge to recognize grammatical sentences in English. Rather, she will go on recognizing them the way she did before she acquired this new knowledge.

(3) Someone who can recognize *X*s can sometimes know and state something interesting about what *X*s are, or about the class of *X*s, or about what all *X*s share. This is a vague characterization and it covers a lot of ground. In the case of grammar, for example, it runs the gamut from being able to specify what Chomskians call a descriptively adequate grammar (see K&S[3]), to being able to state some or all of the rules that might be found in a traditional grammar book for English, to being able to state some even less explicit characterization of the class of English sentences (like those proposed in K&S [3]).

When you talk about "explicit knowledge" of *X*, it seems to be knowledge in this vicinity that you have in mind. But since saying something interesting about what *X*s are is both vague and open ended, we don't think that your notion of explicit knowledge is clear or interesting or useful. Perhaps we're being unfair here. Perhaps you do have some clear idea about what *S* needs to be able to specify about *X*s if *S* is to count as having explicit knowledge about them. Perhaps, that is, you do have a clear answer to Q14, and perhaps that answer makes explicit knowledge an interesting and important notion. But we're skeptical. The ball is in your court here.

At this point, we imagine, you might be thinking: "What's the big mystery? It's all to do with *patterns*. We know that there must be a pattern that unifies all the grammatical sentences of English. And to have explicit knowledge, a person has to be able to (as you say in FJ[3]) '*produce a sentence that gives the pattern*.' To say why we find this unsatisfactory, we'll have to once again take up the question we raised in K&S[3], namely. . . .

2.3 What are patterns? In addressing this question in your most recent letter (FJ[3]) you suggest what may be the nub of the problem, and we think you are spot on.

> [C] Why is it that we seem to so often to puzzle each other? Maybe the nub of it is in your question "*But what could it mean to say that there is a pattern that unifies all the grammatical sentence of English*?"
>
> You go on to discuss this as if it were a question about language. For example, you discuss the idea that it means "*the set of grammatical sentences must be recursive (or, a bit more modestly, that it must be possible to characterize them with a finite set of rules or principles*)." For me it isn't. If there's a pattern, it will in principle be possible to capture it in words and there will be many ways to do this. But the existence of a pattern is a fact about the world—in particular about a certain set of sentences—and not about words. Square things are alike; we can capture the fact in words; but the similarity is not a fact about words. The difference between square things and grammatical sentences is the extent to which the similarity is disjunctive but the sense in which both are cases of things falling under a pattern is the same. (FJ[3])

Here's our diagnosis of the problem: We believe that most of this talk of patterns and similarity and structure and things being alike is (something like) a *metaphor*. If interpreted in a natural (and nonmetaphorical) way, it is simply false that all grammatical sentences are alike, or that they are all similar. Moreover, we do not think there is any good way of unpacking the metaphor. "Philosophy," as Wittgenstein famously proclaimed, "is a battle against the bewitchment of our intelligence by means of language." And we think you've been bewitched.

In an effort to break the spell, we'll set a cluster of examples aimed at making it clear why we don't think the notion of a pattern makes any clear sense when talking about grammatical sentences in English, or in many,

many other cases where you seem to think the pattern metaphor is unproblematic. Each one of our examples will be an artificial language whose sentences are sequences of standard Arabic numerals, of arbitrary length. So, for example, any of the following might be a sentence in one of our languages:

1221
000000500000000999999999999
17
1234567
9082375830275768473658686878945352759

(Of course, nothing turns on calling these "artificial languages" or on calling the numeral sequences "sentences." You can, if you prefer, just think of them as sets of numeral sequences.)

Language 1 is an infinite set of numeral sequences of arbitrary length; the members of the set are *completely random*. (We like to imagine having some device for randomly generating numeral sequences of arbitrary length, and just letting the device run for ever. But that's window dressing.)

Language 2 is a finite language. It is formed from Language 1 by taking the first 1,000 numeral sequences. Those "sentences" and only those are grammatical in Language 2.

Language 3 is another infinite language. It includes '17', '34', '51', and all the other numeral sequences that are (standard names of) multiples of 17.

Language 4 is the union of Language 2 and Language 3.

OK. Now let's ask: As you use the notion of a pattern, in which of these languages is there a pattern to be found? We expect that 1 and 3 are the easy cases. The sentences in 1 were chosen at random, so we assume you want to say there is no pattern there. Indeed, if you think there *is* a pattern there, then your notion of a pattern is vacuous; it excludes nothing. We assume that you would say there is a pattern in Language 3, since the sentences in L3 all have something in common; they are all multiples of 17.

Now let's turn to the hard cases. Is there a pattern in Language 2? Here, we think, different considerations pull in different directions. The sentences in L2 were chosen at random. And that surely suggests that there is no pattern there. On the other hand, it would be possible to learn to recognize sentences in L2 after a finite number of presentations,[8] and in FJ[2] you say: "We know there must be a pattern that unifies the grammatical sentences in English, otherwise we could not acquire the ability to recognise the grammatical sentences in English after a finite number of presentations. Ditto for Russian etc." Well, ditto for Language 2, too. So there "must be a pattern." But now it looks like you are committed to the view that there is a pattern in *any* finite set of objects. So, given the things you say about patterns in quote [C] above, it looks like you are committed to saying that the objects in any finite set are "alike" or "similar" though perhaps the similarity they share is "disjunctive." And if you do say that, then we simply have no idea what you mean by 'alike' and 'similar'.

At this point, you might protest: "Of course there is no pattern in the first 1,000 elements of a set of numeral sequences chosen at random. It is the finiteness of Language 2 that's responsible for the fact that it's possible to learn it after a finite number of presentations, but the learnability test was never meant to apply to finite sets."

But then what will you say about Language 4? Since it includes L2, if there is no pattern in L2 it's hard to see how there could be a pattern in L4. But, like both L2 and L3, it is possible to learn to recognize sentences in L4 after a finite number of presentations, and L4 is infinite. So is there a pattern, or isn't there? We have no idea how to answer this question. And the reason we don't is that your notion of pattern is seriously underspecified. There is, we believe, no clear way to extend the pattern metaphor to interesting cases like language. ("See how high the seas of language run here.")[9]

2.4 How can your account of implicit knowledge be generalized to folk psychology? Before starting on this topic, we'd do well to summarize the state of play. If we've got you right, then when you say that *S* has implicit knowledge of *X* (where *X* can be grammar, wffs, circles, etc.) you are attributing a complex ability or capacity to *S*. The capacity has three features:

(i) an ability to recognize *X*s;
(ii) an ability to say of non-*X*s why they are non-*X*s;
(iii) an inability to say what *X*s are explicitly.

Though we have problems with (ii) and (iii), we take (i) to be relatively unproblematic. And since (i) will presumably be crucial in any application of the notion of implicit knowledge to folk psychology, that's where we propose to focus our attention.

When we started this exchange of letters our goal was to better understand your conception of folk psychology. The way we got involved in the extended discussion of implicit theories was that, in your first letter, you made it clear that you think folk psychology is an implicit theory. You also introduced the analogy with grammar. Here is the relevant passage:

> Defenders of folk psychology often say that it is an implicit theory. I do. But what I mean by this is one, but only one, of the things sometimes meant when it is said that we have an implicit theory of grammar. There is an implicit theory that drives our classifications of sentences in languages we have mastered into the set of the acceptable and the set of the nonacceptable sentences. This is the theory we make explicit by interrogating our intuitive classifications and which, when extracted and recorded in words, makes its way into grammar books as an explicit theory of grammar. That's how grammar books get written. (FJ[1])

Now that we've made some progress in understanding what you mean by 'implicit theory' (or at least we hope we have) we've started to think about what it might mean to say that folk psychology is an implicit theory, focusing on (i), the recognition component of implicit theories. And we need some help. Presumably, on your view, if a person has implicit knowledge of folk psychology, she has a complex ability which includes an ability to recognize *something*. What we are not clear on is *what the something is*. The problem is not that there are no candidates. Quite the opposite, there are lots of candidates, some more problematic than others.

In "What Is Folk Psychology?" Stich and Ravenscroft (1996, 124) assembled "a partial list of the 'folk psychological' capacities or abilities that need explaining." These included the following (the numbering here is not the same as the numbering in the article):

1. The ability to attribute beliefs, desires, emotions, and other mental states to oneself and to others.
2. The ability to predict how people will behave.
3. The ability to construct explanations of people's behavior couched in mentalistic terms.
4. The ability to judge the correctness or incorrectness of psychological principles or generalizations couched in mentalistic terms. (What we had in mind here were Lewis-style "platitudes" like "When a normal person is looking at a traffic light which changes from red to green, she usually comes to believe that it has changed from red to green.")

More recently, Shaun Nichols and Steve published a book, *Mindreading* (Nichols and Stich 2003), in which they pointed to a number of other abilities that the folk have which need explaining. Here are some drawn from the list on page 77:

5. The ability to predict the inferences that other people will draw.

6. The ability to attribute a limited number of perceptual states to other people. (It turns out that there are surprising shortcomings here.)
7. The ability to predict some of the decisions that other people will make. (Here again there are surprising shortcomings.)

Now, if one is not too fussy, one might describe most or all of these as recognition abilities. (1) I can recognize when *S* believes that *p*, and when *S* does not. (2) I can recognize when *S* will call the police and when *S* won't. (4) I can recognize psychological generalizations ("platitudes") that are true (or plausible, or sound right) and generalizations that are not. And so on.

As noted in the Stich and Ravenscroft article, the grammar analogy works best for (4)—judging the correctness of generalizations. Just as our intuition tells us that some sentences are OK and others are crook (as you so delightfully put it), so too our intuition tells us that some putative platitudes are OK and others are crook. So perhaps when you say that people have implicit knowledge of folk psychology, what you mean[10] is just that they have the ability to recognize good and crook putative platitudes, and all the rest of these abilities are irrelevant. On the other hand, perhaps you would want to include some or all of these other abilities among the recognition abilities that a person must have if she has implicit knowledge of folk psychology. There are, we think, some very sticky issues to be faced if you go the latter route. But there is no point in exploring those until we know which way you go here. And besides, this letter is already way too long.

All the best,
Kelby and Steve

FJ[4]: Letter from Frank Jackson [4], August 4, 2005

Dear Steve and Kelby,

The two key things I need to say to assist with our mutual comprehension problem is, first, I do think any finite set of items automatically exemplifies a pattern, and, second, without clause (ii) we would not necessarily have a case of (implicit) knowledge of what it takes to be a wff, say. Let me say something about both points in turn.

1 Patterns

My attitude to patterns is like Quine's to objects. Very inclusive. The city of London plus the coin on my desk is an object. Not one anyone would have much interest in—except to make a point in a philosophy seminar, and not one there is a word for in English though we could introduce one, but an object all the same. Ditto for patterns. Thus any finite sequence of numbers exemplifies at least one pattern. Actually if we are dealing with a sequence of word tokens for numbers like the sequences in your letter, there will be a huge number of patterns: being written on a certain page, being produced by such and such a person, being of so-and-so a temperature at such-and-such a time, etc. Of course the set of English sentences is not finite and the pattern that unites the grammatical ones is reasonably unified and of interest—indeed some importance in everyday life. Ditto for wffs except the interest is more limited.

This means that saying that there exists a pattern in and of itself is typically a very weak claim. What isn't a weak claim or need not be is saying that so-and-so a pattern is what some bit of English picks out, or saying that some pattern or other is something implicitly or explicitly known.

2 Clause (ii)

You can have the ability to classify formulas of logic into wffs and non-wffs in the absence of an ability to give the, or a, general formula that covers all cases without having an implicit theory of wff-ness. This is why clause (ii) is needed. Think of a student a bit like *Clever Hans*. The student correctly classifies formulas into wffs and non-wffs but is doing it—unwittingly let's suppose—by noting something about the expression on the tutor's face when she, the tutor, looks at any given formula. This student does not have implicit knowledge of what it takes to be a wff.

The key point about clause (ii) is that it means that we could construct a theory of grammar or wff-ness in the explicit sense from the knowledge of the person who only has an implicit theory of grammar or wff-ness. Indeed that is how explicit grammars were constructed and how many of us came to have an explicit theory of wff-ness. The case-by-case knowledge is enough to allow us to find the words that cover all the cases.

Hope this helps,
Frank

K&S[5]: Letter from Kelby Mason and Steve Stich [5], September 11, 2005

Hi Frank!

One issue that needs to be addressed is where to end the discussion. In the last section of our letter of K&S[4] we raised some questions about how

your account of implicit theories applies to the case of folk psychology. However, you did not address those questions in FJ[4]. We're inclined to think that the article might be of greater interest if it included some discussion of this issue. So, if you are not yet tired of the exchange, we'd like to encourage you to say something about it.
All the best,
Steve and Kelby

FJ[5]: Letter from Frank Jackson [5], November 14, 2005

Dear Steve and Kelby,
I fear we are still at cross-purposes on many issues. I've tried to be as clear as I can in my response.

1 What Do I Mean by Saying That S has an 'Implicit Theory'?

I mean *S* has a theory which is not explicit. I'm not 'attributing a complex ability or capacity' to *S*, though *S* may have—typically will have—various complex abilities.

What's a theory that is not explicit for *S*? It is one (i) *S* holds but (ii) *S* cannot give the content in words. If *S* is asked on an exam to state the content of theory *T*, when *S* only knows *T* implicitly, *S* fails that question.

To have a theory is to have a certain view about how things are, and so any philosopher who thinks that one can have views about how things are which cannot be put into words should believe in implicit theories.

The material on recognition in my discussions of implicit knowledge and theories is not part of what it is to be implicit. The reason for giving the examples where we can recognize that something is a *T* (to be a *T* is to satisfy *T*) is because they are a good source of cases where subjects have an implicit theory. However, recognitional ability is not part of what it is to have an implicit theory.

Here is an example where a theory is implicit but there is no recognitional ability. Fred is 'motion blind' in the sense that he cannot see things *as* moving. However, whenever given information like:

Mary is in Melbourne at t_1; Mary is in Sydney at t_2; Melbourne and Sydney are in different places

Fred infers that Mary has moved.

However, he cannot produce anything like '*x* moves iff *x* is in different places at different times' when asked for the conditions under which something counts as moving. He has, that is, two deficits by comparison with most of us. He cannot see things as moving, and he cannot give the rubric.

He has (say I but I think this is also what Lewis and many would say) an implicit grasp of what motion is; he has an implicit theory of motion.

2 So How Does My Account of Implicit Knowledge Generalize to Folk Psychology—Your Question?

It depends on what you mean by 'folk psychology'.

(i) If you mean *anything* the folk mostly hold about psychological states, there are many things they believe which are not implicit. Take:

Pain is unpleasant.

This is something the folk believe and they can put it in words easily enough.

(ii) There is, however, a view about the mind associated most especially with Lewis and analytical functionalism more generally (I know there are ways of carving up the territory that excludes Lewis from the class of analytical functionalists but here they belong together) that can be sketched as follows.

There is a theory *P* such that

(1) it is held by the folk;
(2) it is largely implicit (for the folk);
(3) a subject is in mental state *M* iff the subject satisfies the open sentence for *M* that comes from *P* in the usual way.

Those who hold this position often call *P* 'folk psychology'. *P* is *not* everything the folk believe about the mind. It is the theory that satisfies (1) to (3) above, if such there be.

It is on the second sense of 'folk psychology'—the sense on which folk psychology is that which satisfies (1) to (3) above—that the notion of an implicit theory is important. Indeed it is crucial for the plausibility of the view about the mind just outlined. This is because it is implausible that there is an explicit theory that satisfies (1) to (3) above. In other words it is implausible that:

There is a theory *P* such that

(1) it is held by the folk;
(2*) it is explicit (for the folk);

(3) a subject is in mental state *M* iff the subject satisfies the open sentence for *M* that comes from *P* in the usual way.

However, nothing I say about what an implicit theory is *supports* the above view of mind. One might well hold that there are implicit theories while rejecting the Lewisian view of the mind sketched above.

I am wondering if the misunderstanding between us arose because you expected more from the notion of an implicit theory? Perhaps you thought the notion was in itself an argument for the above view of the mind?
Best,
Frank

K&S[6]: Letter from Kelby Mason and Steve Stich [6], April 15, 2006

> Results! Why man, I have gotten a lot of results. I know several thousand things that won't work.
>
> —Thomas Alva Edison, who did most of his work just a few miles from Rutgers

Dear Frank,
Back in July, when we sent off K&S[4], we thought we were making real progress in understanding your view. But your two most recent letters, FJ[4] and FJ[5], make it clear that we were unduly optimistic. We're still almost totally unable to understand your view, and we suspect that you are equally puzzled about why we don't understand you. Since this must be our last letter, what we'll do here is offer our take on the state of play in this dialogue. For the most part, it will be a catalog of things we do not understand. Still, if Edison was right, we've gotten a lot of results, even if the lightbulb has only flickered now and then.

Let's start with a brief reminder of what we *thought* we understood in K&S[4]. At the core of our mutual miscomprehension is your notion of implicit knowledge, and in K&S[4] we proposed an analysis of that notion. Implicit knowledge of *X*s (where *X* can be wffs, grammatical sentences in English, circles, or any other "pattern") consists of two abilities and an inability:

(i) the ability to recognize *X*s
(ii) the ability to say for each non-*X* why it is not an *X*
(iii) lack of the ability to "produce an open sentence . . . that gives the pattern" FJ[2]

We went on to explain that we had concerns about (ii) and (iii), but that (i), at least, was more than clear enough to work with.

The most discouraging part of FJ[5], for us, is your insistence that we were wrong about (i). "[R]ecognitional ability," you write, "is not part of what it is to have an implicit theory." The context makes it clear that recognitional ability is not *necessary* for implicit knowledge, on your account; it is less clear whether you would say that conjoined with (ii) and (iii) it is not *sufficient*.

We suspect—and hope—that here we're just talking past one another again. As we view the example you give, it sounds reasonable to say that motion-blind Fred *does* have a recognitional ability, albeit a highly inferential one. Given the relevant facts about Mary, he can recognize the case as one of motion. But it seems that you consider this as not recognitional, because Fred can't *perceive* the motion. So perhaps you are interpreting "recognitional ability" as a kind of perceptual ability. If so, then perhaps you'd want to say that Fred has a *classificatory* ability but not a *recognitional* one. So let's try rephrasing (i) thus:

(i′) the ability to *recognize or classify X*s (as *X*s)

We have our fingers crossed that you'd be willing to grant this kind of ability as a necessary part of implicit knowledge. If not, then we're back to square one. If (i′) (or something in the vicinity) is not even *part* of what it is to have an implicit theory, then we have, near enough, no understanding at all of what you think implicit knowledge is.

In the same section of FJ[5] in which you say that recognitional ability is not part of implicit knowledge, you make one more attempt to say what you mean when you say that *S* has an implicit theory. "I mean *S* has a theory which is not explicit. . . ." And "[t]o have a theory is to have a certain view about how things are, and so any philosopher who thinks that one can have views about how things are which cannot be put into words should believe in implicit theories." One of the disadvantages of having this discussion via email rather than across a table is that it is sometimes hard to discern the "tone of voice" with which comments are made. The first time we read this, we imagined it written in exasperation. Surely you couldn't think this would help us to understand your view, so you were probably just throwing up your hands and giving up on us. We still, mostly, think this is the way to read the remarks we've quoted. But from time to time we entertain the idea that you *did* think this might help because in your corner of the philosophical world "having a certain view about how things are" is an expression which is clear enough, and precise enough, to use in careful philosophical discussion. If so, then the divide between your corner of the philosophical world and ours is even greater than we had imagined. *Of course* we think that "one can have a view about how things are." But in our philosophical dialect, the expression is far too vague to be of any help in understanding what you mean by 'implicit theory'.

Let's turn, now, from (i) to (ii). In K&S[4] we went on at some length about why we found your condition (ii) puzzling, and in FJ[5] there were two short paragraphs in which you expanded on your view. Unfortunately, we did not find them very helpful. Indeed, as far as we can see, they did not really address our concerns at all. Here's how we see the state of play on (ii). You have articulated the requirement in several different ways. In FJ[2], where the example in question is implicit knowledge of grammatical sentences in English, you say: "I can say case by case for each crook sentence *how to fix it*" (emphasis ours). In FJ[3], where the example at hand is implicit knowledge of wffs, you say that those who have implicit knowledge "know case by case why a non-wff is a non-wff." In both K&S[3] and K&S[4] we asked for clarification on how to interpret these remarks.

The problem, as we see it, is that on the natural interpretation it is simply *false* that the sorts of people to whom you would clearly attribute implicit knowledge of *X*s know why non-*X*s are non-*X*s. Let's focus on grammar, which has been center stage in much of our correspondence. Both of us are native English speakers, and it is clear that you would attribute to us an implicit knowledge of grammaticality in English (or an implicit theory of English grammar). Both of us recognize that the sequence of words 'would brick that clear is and bit us to' is not grammatical in English. But if asked *why* it is not grammatical in English, we would have no idea what to say. Nor would we know what to do if you asked us to "fix it." In FJ[3] you say this is not true; you insist that we *do* know how to fix sentences like this. "Here's one way to fix it. Put quote marks around it and add 'is not a sentence of English' to the result." But as we note in K&S[4], this makes (ii) an *extremely weak* requirement. You are of course, free to interpret the requirement in any way you wish. But in several places, including most recently FJ[4], you have maintained that the sort of "case by case" knowledge required by (ii) is going to do important work for us. Your most explicit statement on this is the following in FJ[2]: "I can say case by case for each crook sentence how to fix it (there will be many ways of course) *and that case-by-case information is enough to construct in principle the sentential representation*" (emphasis added). But on the weak interpretation of (ii) suggested by your "put quote marks around it . . ." proposal, this strikes us as patently absurd. How could the information gleaned from *that* sort of "fixing" be of any use at all?

So as we see it, here's the bottom line on (ii). Interpreted in what we take to be the natural way, it is just false that speakers typically know how to fix nonsentences or to say why they are not grammatical. So on this interpretation, speakers typically fail to satisfy (ii) and thus they do not have implicit knowledge of the grammar of their language. And that's bad news for you. Interpreted in the weak "put quote marks around it" way, the information provided by speakers' case by case "fixes" is utterly useless. And once again, that's bad news for you. Clearly, you've got some work to do here. You need to say more explicitly what (ii) requires, and to explain why you think the case-by-case information that can be obtained from people who meet that requirement is of any use.

Let's now move on to (iii). In K&S[2] we asked what we thought was a straightforward question:

Q14 What are the constraints on the "belief capturing" open sentence that must be available to a person if his theory is to count as explicit rather than implicit?

And in S&K[4] we protest that "you never answered this question. The closest you come is to talk about a sentence that 'gives the pattern.'" We went on to say why we thought this was too vague to be of help:

> Someone who can recognize *X*s can sometimes know and state something interesting about what *X*s are, or about the class of *X*s, or about what all *X*s share. This is a vague characterization and it covers a lot of ground. In the case of grammar, for example, it runs the gamut from being able to specify what Chomskians call a descriptively adequate grammar (see S&K[3]), to being able to state some or all of the rules that might be found in a traditional grammar book for English, to being able to state some even less explicit characterization of the class of English sentences (like those proposed in S&K[3]).
>
> When you talk about "explicit knowledge" of *X*, it seems to be knowledge in this vicinity that you have in mind. But since saying something interesting about what *X*s are is both vague and open ended, we don't think that your notion of explicit knowledge is clear or interesting or useful.

In FJ[5] you take up the matter again: "What's a theory that is not explicit for *S*? It is one (i) *S* holds but (ii) *S* cannot give the content in words. *If S is asked on an exam to state the content of theory T, when S only knows T implicitly, S fails that question*" (emphasis added). As we said earlier, in discussing the first part of this quote, we're not sure what "tone of voice" to attribute to you here. Perhaps you are so exasperated by our questions that you can no longer take them seriously. Perhaps you mean the emphasized passage as a joke. If it is intended as a joke, let us respond in kind.

Standards on exams differ dramatically from place to place and from time to time, and in the United States, "grade inflation" has become rampant during the last few decades. Thus at many U.S. schools nothing would count as implicit knowledge, by your standards, because *no one fails*! It's not a great joke, to put it mildly. But it does have a point. Your failing-the-exam account of implicitness is far too vague to be of help to us in trying to understand your view.

Finally, let us say something about the vexed topic of patterns. It was here that FJ[4] held its biggest surprise. Frankly, we never entertained the possibility that you would say that "any finite set of items automatically exemplifies a pattern." And we are still struggling to come to grips with the implications of this revelation. Why were we so surprised? Well, in talking about patterns, both in this correspondence and elsewhere, you repeatedly use words like "similarity" and "alike" and "structure," and you use squares and circles as standard examples of things that exhibit a pattern. Here is an entirely representative example from FJ[3]: "[T]he existence of a pattern is a fact about the world. . . . Square things are alike; we can capture the fact in words; but the similarity is not a fact about words. The difference between square things and grammatical sentences is the extent to which the similarity is disjunctive but the sense in which both are cases of things falling under a pattern is the same."

Once you have made it explicit that, on your conception of pattern, *any* finite set of items exemplifies a pattern, all of this is, at best, seriously misleading. To emphasize the point, consider an example. Let *R* be a finite set of things chosen completely at random. For concreteness, imagine that they are chosen as follows. We launch a military drone on a series of flights around the world. Every second it snaps a photo, and analysts identify the most salient object in the center of the photo. Then a truly random physical process—one that depends on cosmic ray impacts, perhaps—is used to decide, for each salient object, whether or not it is in *R*. When *R* contains exactly 982,475,893 members, the flights are terminated. On your account, the members of *R* exemplify a pattern. So readers of the above passage are encouraged to think that the members of *R*, like square things, "are alike"—that there is some disjunctive "similarity" between them, and that the existence of the *R* "pattern," and of the "pattern" associated with *every other finite random set*, is a "fact about the world." Is it any wonder that we were misled? Can there be any doubt that most other readers would have been misled as well?

You acknowledge in FJ[4], that your "attitude to patterns is . . . [v]ery inclusive." And we have been lamenting the fact that you were not more explicit about this earlier in our correspondence and in your other writings. But now, unless we have misunderstood you yet again, your usage is clearer. For you, the members of *every* set fall under or exemplify a pattern, with the exception of *infinite random* sets. We are still not entirely clear about your conception of patterns, however. In FJ[4] you say that some finite sets, like our Language 2 (in K&S[4]), exemplify "a huge number of patterns." If that's not a slip, then some further clarification is needed.[11]

What puzzles us now is why you think this *very* inclusive notion of pattern is going to be of any interest in the philosophy of psychology or the philosophy of mind. Consider, for example, the sort of classificatory behavior that looms large in your account of conceptual analysis.[12] Is it of any interest to know that an agent's classificatory behavior—or that all the things she classifies as *X*s—exhibit a pattern? As far as we can see, the answer is no, since we know in advance that this has to be the case. For suppose that someone actively *tried* to classify things as *X*s in a way that *X*s exhibited no pattern at all. She is bound to fail, since sooner or later she will die and the set of things she's classified as *X*s, whatever they are, will exhibit a pattern. Indeed, if your comment quoted at the end of the previous paragraph was not a slip, the things she's picked out will exhibit a huge number of patterns. How these facts could be of interest to philosophers of mind or philosophers of psychology, or anyone else, remains a mystery.

We started this letter with Edison, so let's finish with some words of wisdom from another great American inventor, the well-known epistemologist Donald Rumsfeld: "[T]here are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know." We set out on this correspondence in the hope of better understanding your view, thereby expanding our own stock of known knowns. The notions of implicit theory and folk theory have been important in many philosophical debates over the last fifty years, and in much of your own work especially, so wouldn't it useful if we could get clear about how you understand these notions?

So we thought, and so we still think. Alas, we're a long way from there, even after our many exchanges. But we have managed one thing, and that's to reduce our stock of unknown unknowns. At least now we *know* there's lots about your view that we don't know, and don't understand. Which is some kind of progress, at least; even Rumsfeld might approve.

All the best,
Kelby and Steve

FJ[6]: Letter from Frank Jackson [6], November 6, 2006

Dear Steve and Kelby,
I too am sorry we seemed so often to be at cross-purposes but I doubt if giving a detailed set of responses to your summing up of 15 April 2006 (K&S[6]) would assist. Indeed, induction suggests the opposite. However, maybe two general remarks from me will be of value—perhaps for the three of us, or perhaps and hopefully for readers of the correspondence.

1. I want to say something fairly limited about the notion of an implicit theory; you were always reading me as saying something more far reaching. I hold that people can have a theory that things are thus and so without being able to put it into words. If I am right, we need a name for this kind of case and I think 'implicit theory' is not a bad name. I guess I still don't know where you stand on the possibility.

I gave a number of examples where it seems to me that people hold a theory that things are thus and so without being able to put the theory into words. Each example had one or another special feature, as is the way with *examples,* and then—somehow—the discussion became one about those special features.

Take the well-worn example of students who can reliably recognize wffs in logic but cannot give a definition of a wff. When they say that some formula is a wff, surely there is a sense in which they are making a claim about how things are and so have theory about how things are which they are expressing when they that some formula is a wff. At the same time, they cannot produce something of the form: *x* is a wff if *x* is. . . . Well that isn't quite right. They can no doubt produce sentences like: '*x* is a wff iff *x* is a wff', and '*x* is s wff iff *x* is said to be a wff by an expert logic instructor'. What they cannot do is give an illuminating account of what makes a wff a wff. We need a way of describing this kind of situation. I say that they have an implicit theory of wff-hood.

One thing that baffled me about our correspondence is that at the end of it I still did know where you stood on this question. I knew you had worries about the notion of recognition and of the use of 'theory' but not how you yourselves would describe the kind of situation just sketched.

2. I want to distinguish a theory from the statement of a theory. Of course one says things like '$E = mc^2$' is part of STR, which might suggest that the statement itself is part of the theory. But, in my view, it is how things have to be for the statement to be true that is the theory, or the relevant part of the theory. I felt at a number of points in our exchanges that you were using 'theory' more for one or another statement of a theory,

than for a theory in the sense I had in mind. Of course one is free to use 'theory' for a statement-like animal, but then you would need another term, say, 'theory*', to discuss what I was discussing.

If a theory is not a statement, what kind of animal is it? I think it is a way things might be, and a true theory is a way things might be that obtains. '$E = mc^2$' contends that a certain way things might be obtains at the actual world. (What about inconsistent theories and mathematical theories—good questions for another time.) It is, that is, a view about what our world is like. This is the same as saying that it is a view about the patterns exemplified in the actual world.

You worry that I have a very inclusive notion of a pattern. But I don't have a very inclusive notion of a pattern of *interest*, or a pattern *affirmed by one or another theory*, or . . . , so I was at a loss to understand what worried you so much about my inclusive notion of a pattern per se.

Best, as always,
Frank

Notes

1. Stich and Ravenscroft 1996.

2. Weinberg, Nichols, and Stich 2001; Nichols, Stich, and Weinberg 2003; Machery et al. 2004.

3. See, e.g., Gopnik and Meltzoff 1997 and the critique in Stich and Nichols 1998.

4. See, e.g., Stich 1993.

5. Chomsky also required that these rules should entail many facts about the grammatical properties and relations of sentences in the language. But for present purposes we can ignore these complications.

6. It is the account developed in Stich and Ravenscroft 1996, sections 2 and 3.

7. Suppose there is someone who fails to meet the able-to-fix-it requirement. All we need do is teach him the put-it-in-quotes-and-append–"is not grammatical in English"-trick, and voilá, he can fix *any* ungrammatical sentence.

8. If you doubt this, let L2 contain the first 100 sentences in L1, rather than the first 1,000.

9. Note that although our example of L4 is both artificial and quite simple, it does have some important features in common with natural languages. Like L4, natural languages can be described by a bunch of recursive rules plus a finite bunch of apparently random exceptions and special cases and lexical rules. Pattern? No pattern? We have no idea, because we don't know what is being asked.

10. Or, better, the recognition component of what you mean.

11. Do *all* finite sets exemplify a huge number of patterns? How about nonrandom infinite sets? Do they, too, exemplify a huge number of patterns? And just *how* huge is that huge number? Do some (or all) finite sets and nonrandom infinite sets exhibit an *infinite* number of patterns?

12. In Jackson 1998a, chapter 2.