CHAPTER SIX

PSYCHOLOGICAL CONCEPT ACQUISITION

THEODORE BACH
BOWLING GREEN STATE UNIVERSITY, FIRELANDS COLLEGE

1. Introduction

A distinguishing factor among theories of folk-psychology is their explanation of the nature and development of psychological concepts, particularly the belief concept. The goal of this chapter is to offer novel and critical evaluations of the simulationist, modularist, and theory-theorist explanations of psychological concept acquisition. The chapter unfolds as follows. Section (2) discusses Gordon’s simulationist construal of psychological concept acquisition and criticizes Gordon’s use of the notion “mental location.” Section (3) critically evaluates Scholl and Leslie’s modularity account of psychological concepts. Section (4) defends Bartsch & Wellman’s scientific theory-theory (STT) view of psychological concept acquisition against criticisms from Nichols & Stich (2003) and Goldman (2001), and section (4.1) argues that Bartsch & Wellman’s account must be supplemented with a theory of hypothesis discovery.


Both defenders and critics of simulation-theory generally agree that mental simulation does not underlie the development of propositional attitude concepts such as “belief” and “desire.” If simulation requires projecting a psychological state from self to other, and if projection requires first categorizing one’s internal mental state, then it would be circular to explain grasp of attitudes via simulation. Using this line of reasoning,
many simulationists endorse a functionalist account of psychological concepts, an introspectionist account, or some combination.¹

Robert Gordon, on the other hand, claims that the explanatory potential of simulation with respect to psychological concept acquisition has been overlooked because researchers falsely assume that simulation requires the analogical projection of mental states (what Gordon calls “transference”). Gordon recommends that we instead construe simulation as involving “transformation”—a process through which the simulator “recentres” their “egocentric map” such that the pronoun ‘I’ refers to the simulatee rather than the simulator (Gordon 1995, 55–56).² In so far as the self transforms into the other during simulation there is no need to analogically project mental states from self to other, in which case there is no need to invoke psychological concepts in the explanation of simulation.

The idea of transference helps avoid a circular account of psychological concepts, but it does not tell us about the nature of such concepts. To develop a positive account of psychological concepts Gordon takes up Gareth Evans’ notion of “ascent routines.” Ascent routines involve answering a meta-cognitive question (“Do you believe that tomatoes are ripe in August?”) by descending one semantic level in order to answer it as an object-question (“Are tomatoes ripe in August?”), and then ascending back to the meta-cognitive level with the answer. The “I believe …” preface, it would seem, is performative. Ascent routines shed light on psychological concept acquisition when viewed in the context of transformative simulation:

“… in a simulation of O … ‘I’ refers exclusively to O, the individual on whom my egocentric map has been recentered. So I settle the question of whether O believes that p simply by asking, within the context of a simulation of O, whether it is the case that p.” (Gordon 1996, 60)

But as Gordon is aware, this ability to execute ascent routines falls short of psychological concept mastery because it “does not enable [children] to make genuine, comprehending ascriptions of belief” (Gordon 1996, 15). Gordon claims that genuine and comprehending belief ascriptions require a grasp that facts have a “mental location,” that is, that facts are facts to some individual:

¹ Goldman (2001) advances a “dual representational” account of propositional attitude concepts that features both a theoretical component and an introspectionist component.
² Gordon illustrates this idea in reference to method-acting. Allegedly, method actors do not pretend at being others but actually become others. Folk-psychologists are everyday if less sophisticated method actors.
[Children] fail to grasp several components of the concept of belief, but the one that is paramount, because it is presupposed by all the others, is the general idea that a fact (about Mickey Mouse, for example) can have a mental location. (Gordon 1996, 16).

According to Gordon, then, grasping mental location is the key conceptual achievement involved in grasping belief (Gordon claims that children learn the concept of mental location through practice with mental simulation).

I agree with Gordon that grasping a psychological concept requires grasping what is “paramount” to that concept. I disagree, however, that mental location is that paramount feature. All propositional attitudes—desires, wonderings, doubtings, etc.—have a mental location. As a result, mental location fails to distinguish among the various propositional attitude concepts. Instead of understanding mental location as the essential feature of belief, I suggest that we understand it as a description of, or metaphor for, perspectival subjectivity. If this is correct, then the “radicalness” of Gordon’s theory reduces to the less controversial claim that grasp of the attitudes presupposes knowledge of perspectival subjectivity, and that knowledge of perspectival subjectivity is acquired through simulation. As the account stands, then, it fails to inform us how children learn specific propositional attitude concepts such as the belief concept and the desire concept.


Modularity approaches to folk-psychology (Fodor 1987; Baron-Cohen 1995; Scholl and Leslie 1999) view psychological concepts as theoretical knowledge structures that have an innate basis and which satisfy a sufficient number of the Fodorian conditions for modularity.\(^3\) The *locus classicus* for this view is Baron-Cohen (1995) which describes a set of four modules—the intentionality detector, the eye detection system, the shared attention mechanism, and the theory of mind module (ToMM)—that collectively comprise mind-reading ability. Alan Leslie and colleagues have developed the modular view in detail. On both accounts, the belief concept is a built-in feature of ToMM.

---

\(^3\) Baron-Cohen (1995) lists a version of the conditions on page 57. Scholl and Leslie (1999) list a version on pages 133–134. Baron-Cohen’s conditions are: (1) domain specificity, (2) encapsulation, (3) obligatory firing, (4) shallow outputs, (5) speed, (6) inaccessibility to consciousness, (7) a characteristic ontogenetic course, (8) a dedicated neural architecture, and (9) a characteristic pattern of breakdown.
In a number of passages, Scholl & Leslie (1999) emphasize that ToMM has an innate basis. Because ToMM encodes the belief concept, this is the claim that the concept belief is innate. At first glance this view fits awkwardly with the gradual and seemingly rational development of the mentalizing skill. Scholl and Leslie explain the gradual development of theory-of-mind (ToM) by appealing to developmental changes in the non-modular “selection processor.” The belief concept is there all along, but the false-belief task exploits the absence of the selection-processor in order to create performance errors. Moreover, Scholl and Leslie claim that their nativism predicts a particular type of developmental trajectory and that this prediction is borne out by the empirical evidence. For the remainder of this section I will focus on this second claim.

Scholl & Leslie argue that a modular account better predicts the cultural universality and uniformity of ToM than does the STT account. On the modular account the essential character of ToM is innately fixed: one’s environment should not significantly affect the character of ToM but only the timing of its expression. This account predicts that people in different environments will acquire the same basic ToM though perhaps at different rates. This is indeed the case. Domain-general learning accounts, on the other hand, explain ToM as a learning response to the local social environment—change the environment and you change the theory. According to domain-general accounts, the universality of ToM results from the universality of basic social and psychological phenomena. The two theories appear empirically equivalent. However, as Gopnik notes, a forbidden experiment could differentiates the empirical commitments of the theories:

… there is a simple crucial experiment that could adjudicate between the theory-formation theory and a modularity theory. Unfortunately, like nearly all the crucial experiments in psychology, it is immoral. Put the child in a situation that was the folk-psychological equivalent of someone acquiring a Pidgin language. That is, arrange for her to be surrounded and cared for by creatures with an equally complex but radically different psychological organization than our own. Other things equal, the modularity view would predict that she would end up (eventually) with our folk-psychology. The theory-formation theory would predict that she would instead end up (eventually) with a view of the mind that was at least approximately true of the creatures around her” (Gopnik 1996, 175–176).

---

4 “Our application of the notion of modularity to the domain of ToM results in the following claim: ToM has a specific innate basis” (Scholl & Leslie 1999, 134).
Scholl & Leslie agree that this would be a “crucial and telling experiment” (fn. 7, 144).

Unfortunately, the debate is more empirically intractable than Gopnik, Scholl, and Leslie realize. Scholl & Leslie’s modularity thesis targets only “core” ToM—the basic psychological concepts belief and desire—concepts which can “eventually be recruited by higher cognitive processes” (Scholl & Leslie 1999, 140). But this means that in order to isolate experimentally the domain-general learnability of core ToM along the lines of Gopnik’s suggestion, we would need to secure a radically foreign and “complex psychological organization” that did not contain beliefs and desires. This is likely impossible. Complex psychological organizations may be radically different in their higher cognitive processes, but all such organizations may require, necessarily, possession of something like “primary reasons” (Davidson 1963), actions guided by a belief and desire pair. The necessity here may be conceptual (both Dennett and Davidson seem to argue at various times that we would not call creatures “psychological” if they did not have beliefs). Or the necessity may be nomological if complex behaviour presupposes, as a matter of natural law, a causal relationship between belief and desire structures as these structures are described in current scientific psychology.

This empirical stalemate puts additional pressure on the more general modularist strategy of arguing through negative claims. It is fruitful to view modularity theorists as challenging advocates of STT to provide (a) an empirically supported account of developmental mechanisms that explains children’s theoretical transition as measured by the false-belief task, and (b) an account of developmental mechanisms that predicts the cultural universality and uniformity of ToM. It would be a blow to modularity accounts if domain-general accounts could meet these challenges. Below I defend Bartsch and Wellman’s STT account against some objections and then suggest that this account might meet the modularist challenges if it is supplemented by a theory of hypothesis discovery.

4. Scientific Theory-Theory and Psychological Concepts

Standard versions of scientific theory-theory are “scientific” in two respects. First, children are described as using law-like knowledge structures that appeal to unobservables in the explanation and prediction of cognitive-behavioural phenomena. Second, children’s theories are purported to have the same dynamic features as actual scientific theories. Specifically, children’s theories are constructed through domain-general
learning mechanisms and undergo a “rational” process of conceptual change.

According to Bartsch & Wellman’s (1995) STT account, children progress from a “desire psychology,” then to a “desire-belief psychology,” and finally to a “belief-desire psychology.” Let’s examine these in order. By the age of two, children are desire-psychologists and understand basic emotional and perceptual mental states. Bartsch & Wellman’s account of desire-psychology encompasses three key claims, all of them controversial. One, children have knowledge of desire but not belief. Call this the “desire before belief” thesis. Second, children understand desire as a subjective, internal state but they do not understand desire in representational terms. Call this the “non-representational understanding of desire” thesis. Third, children are able to use their non-representational concept of desire in a limited way in order to predict and explain other’s behaviour. They understand that people act so as to satisfy their desires. Call this the “desire-action schema” thesis. The primary source of evidence for these claims is natural language data on children’s spontaneous mentalistic talk.\(^5\) While researchers agree that Bartsch & Wellman’s evidence is compelling (e.g., Harris 1996, 201), they fault one or several of these claims on conceptual grounds.

It is important to be clear on the difference between a representational and non-representational desire concept. A representational concept posits a representational content as a cognitive intermediary between agent and desired object; the agent is conceived of as wanting the desire-content as expressed in the representation. A representational understanding of “Susan desires a tomato,” for example, will ascribe to Susan an internal representation with the content “Susan eating a tomato” and then a desiring attitude toward that. A non-representational understanding, on the other hand, ascribes to Susan an internal drive for an actual tomato. Here the content of the desire is an actual tomato, or better, the actual tomato exhausts the expression of the content of the desire (the difference is analogous to the difference in the philosophy of perception between direct or “naïve” realism and indirect or “representative” realism).

Now, a \textit{prima-facie} problem for this view is that two year olds (desire psychologists) understand that desires can be unsatisfied and that people desire things that are not present. Even more problematic perhaps is that two year olds understand that desires can be for future states of affairs, e.g. that someone can desire that an empty glass be filled with milk. This

\(^5\) Specifically, Bartsch & Wellman scoured CHILDES, a massive database of transcriptions of children’s talk, and coded every reference to a mental state (e.g., want, think, know, believe, and wonder).
knowledge would seem to require imputing a representation of a filled glass of milk and a desire for the content expressed in that representation. Bartsch & Wellman disagree, using the resource of children’s non-psychological knowledge. Children’s everyday knowledge includes that real objects can be elsewhere and that events are in the future. A non-representational understanding of desire for these things requires that the two year old’s own representational system token a representation of the state-of-affair, and then for the two year old to construe the other (or themself) as wanting that real thing (note the similarity to simulation-theory here). At no point is it necessary to construe the other as entertaining a representational content of the state-of-affair.6

As I framed the dialectical situation earlier, STT needs to demonstrate a rational progression of psychological theories. The progression from a non-representational to a representational concept is pivotal, so opponents of STT would do well to challenge whether there is such a progression.7 Nichols & Stich (2003) and Goldman (2001) do precisely this, claiming that children never possess a non-representational concept of desire. However, I argue that these authors’ criticisms—each aimed at the plausibility of conceptualizing a non-representational desire for future or non-present real world objects—are uncharitable and unsuccessful. Goldman writes:

Suppose the 2-year-old attributes to the target a desire to drink a cup of milk. It does not help to be told that the attributor knows that there have been various milk-drinkings in the past. The question remains: Which milk drinkings in the world does the attributor take the target’s desire to be related to? (Assume, as before, that the desire does not get fulfilled). No answer seems to be forthcoming (Goldman 2001, 211).

6 It is important to distinguish between having a representational concept of desire and using that concept. Adults have the concept but may not always or even often use it. That is, it may be more expedient for them to use the Gibsonian, direct-desire concept of the two year old. However, there are moments when accurate explanation and prediction will demand use of the representational concept, and here, as we will see later, the adult has an important advantage over the two year old.

7 Nichols & Stich are particularly clear on this point. After claiming to have rejected the claim that children progress from a non-representational to a representational concept they write: “And since this alleged conceptual replacement is the central element in the argument that the leap in mindreading abilities during this period is a ‘theoretical development’ subserved by much the same processes that underlie theory change in science, we see no reason to accept this ‘theory revision’ account of the difference between early and later mindreading abilities” (Nichols & Stich 2003, 114–115).
Ironically, Goldman has not taken into account the simulationist flavour of Bartsch & Wellman’s explanation of how the desire-psychologist understands (non-representational) desires for future states-of-affairs. The idea is that the mentalizing two year old knows that cups can be filled with milk and that there is milk somewhere. The child themself represents such milk, that is, they think about an actual glass of milk and then attributes to the target a desire for that actual milk that they are now thinking of. She does not realize that they themselves are entertaining a “representational content”—they “see through” the representational content to the actual object and attributes to the target a desire for that actual object. As such, Goldman’s reading of Bartsch and Wellman seems uncharitable on this point.

Nichols & Stich (2003, 114) claim that on Bartsch & Wellman’s view, in order for a two year old to understand that someone desires a non-present state of affair the child must think that this state of affair is possible. They then argue that two year olds impute to others desires for things that they know to be impossible. Discussing a purported example they write:

At two years nine months, the following exchange occurs between Abe and his mother:

ABE: I want to read the funny papers now.
MOTHER: Well its not Sunday morning is it Abe? You have to be patient and wait for Sunday morning to get here.
ABE: I want the sun come up right now and then the funny papers come.
Abe presumably knows that what he wants is not possible. (Nichols & Stich 2003, 114).

If conceiving of non-representational desires requires conceiving of possible real world-relata, and Abe knows that that the real-world relata is not possible, then we can infer via modus tollens that Abe is not operating with a non-representational concept. I submit that the minor premise is false. Nichols & Stich say in a footnote (fn. 30, 114) that there are various philosophical interpretations of possibility—nomological, metaphysical, and logical. This is of course true, and these modalities apply, generally speaking, mind-independently. That is, it is a mind-independent fact that it is metaphysically and logically but not nomologically possible for the sun (and more controversially Sunday) to arrive early. But let’s grant to Nichols & Stich that the weakest modality—nomological possibility—is what is at stake. Even still, it would be a confusion to say that the major

---

8 Strictly speaking, there are various ways that the sun could appear to rise which do not require violations of laws of nature.
premise of the argument requires real-world relata to be “mind-independently” nomologically possible. Rather, it is the “desire-psychologist” that must take the real-world relata to be a nomological possibility. Now, it is very easy for philosophers trained in modal distinctions to report that “Abe presumably knows that what he wants is not possible”! Rather, at two years and nine months children likely have a very tenuous grasp on what is nomologically possible. At least, the burden of proof is on Nichols & Stich to demonstrate otherwise. I conclude that neither Nichols & Stich nor Goldman offer good reasons to reject the non-representational concept of desire thesis. By their own interpretation of the centrality of this thesis for STT, this is a vindication of STT. I now examine the other stages of children’s developing psychological theory.

At around the age of three children become “desire-belief psychologists.” Here is the central claim:

Our data reveal that children understand the presence of beliefs in the human mind before they grant belief a central role in organizing and explaining mind and action. Hence, in advance of achieving a belief-desire psychology children utilize a desire-belief psychology. We term it desire-belief psychology in order to describe a phase, prototypic of young three-year-olds in our data, in which children recognize the existence of beliefs, at least at times, yet reason about and explain action primarily in terms of desires (Bartsch and Wellman 1995, 149–150).

What the desire-belief psychologist knows, but the desire psychologist does not know, is that there are cognitive states with representational content. That is, they learn that objects, events, state-of-affairs, etc., are represented in the minds of individuals. They begin to talk about these representational states, not to explain behaviour, but to make public their presence. It may be objected that “beliefs” are things that play a particular causal role with respect to behaviour and other cognitive states, and to the degree that a child does not grasp this causal structure they do not have the concept of belief. But this is a point about terminology. This child does not yet have that concept, but they do have a representational theory of mind and refer to internal representations as “beliefs.” In fact, there is strong evidence that children at this age understand that beliefs can be false. Consider two examples of what Bartsch & Wellman term “false-belief contrastives”:

ABE (3;6) The people thought Dracula was mean. But he was nice. (53)
ADULT: I thought it was a bus.
ADAM (3;3): It’s a bus. I thought a taxi.
Other contrastives demonstrate an appreciation between beliefs and imaginings, and beliefs and facts (recall Gordon's explanation of this last distinction). But while desire-belief psychologists grasp that there are internal representations whose contents can be false, they are not yet able to use these representations in the explanation and prediction of behaviour. In other words, they are restricted to the desire-action schema (see Table 6-1).

### Table 6-1. Total Explanations of Action (Bartsch & Wellman 1995, 119)

<table>
<thead>
<tr>
<th>Child</th>
<th>2 Years</th>
<th>3 Years</th>
<th>4 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Desire</td>
<td>Belief</td>
<td>Desire</td>
</tr>
<tr>
<td>Adam</td>
<td>2;6</td>
<td>3;6</td>
<td>5</td>
</tr>
<tr>
<td>Abe</td>
<td>2;5</td>
<td>2;11</td>
<td>23</td>
</tr>
<tr>
<td>Sarah</td>
<td>2;10</td>
<td>4;10</td>
<td>6</td>
</tr>
<tr>
<td>Ross</td>
<td>2;6</td>
<td>3;7</td>
<td>24</td>
</tr>
<tr>
<td>Others</td>
<td>2;1</td>
<td>3;1</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>85</td>
<td>18</td>
</tr>
</tbody>
</table>

As a result, desire-belief psychologists perform poorly on the false belief task. This should not surprise—the false-belief task isolates the causal role of belief in the production of behaviour, so children who rely on the desire-action schema are liable to fail.

Between the ages of three and a half and four, children acquire a belief-desire psychology. At this point they have grasp of the belief-desire action schema which states that “people act on the basis of their beliefs in order to satisfy their desires.” At this point they also pass the false belief task.9

### 4.1 STT’s Account of Transitional Mechanisms.

I agree with Bartsch & Wellman’s description of the stages of psychological theory and I have defended this view against several objections. I am somewhat less sympathetic to their account of transitional mechanisms, not because that account is mistaken, but because it is incomplete and at times uninformative. Their account is that when desire

---

9 Some recent studies (e.g. Onishi & Baillargeon, 2005) challenge STT’s developmental time-table by purporting that young infants have an implicit grasp of false-belief. However, the import of these studies remains unclear. Several researchers (e.g. Ruffman & Perner 2005) argue that the inference from the data to the conclusion about false-belief awareness is not warranted.
psychologists encounter phenomena not easily explained in reference to a non-representational concept they will generate “auxiliary hypotheses” which describe states with representational content that can do the explanatory work. Similarly, when desire-belief psychologists encounter behaviours that are not explained by the desire-action schema they will generate auxiliary hypotheses which invoke the explanatory importance of belief-states. While such hypotheses are at first auxiliary, they are eventually incorporated into the child’s more general theory-of-mind and given an important role:

In the process of utilizing desire-psychology to predict and explain people’s actions and emotional reactions, children may encounter certain puzzles or explanatory failures. For example, the child may notice that he and another person have similar desires but engage in two different actions…At first, an appeal to representational states is recruited only when desire psychology breaks down or in order to provide a mental state explanation that is consistent and relevant. Consideration of what the actor wants still drives children’s explanatory efforts…Consideration of beliefs as framing the actor’s desires, not merely existing as auxiliary to desire psychology, is the key notion behind belief-desire reasoning…What began as an auxiliary hypothesis becomes increasingly typical and finally prototypical of all mental states (Bartsch & Wellman 1995, 171–172).

However, even if this account is correct, it fails to inform how children make the relevant theoretical transitions. Specifically, while it tells us which explanatory contexts require auxiliary hypotheses, it does not shed light on how children actually develop or discover such hypotheses. These moments of epistemic discovery are precisely those that nativists such as Fodor have argued cannot be explained on domain-general accounts of conceptual change.

5. Conclusion

Researchers disagree about the nature and development of psychological concepts. This essay provided new reasons to be sceptical about simulationist and modularist proposals. I then defended the scientific-theory-theory account of psychological concepts against several objections, while also claiming that this account fails to provide an adequate explanation of transitional mechanisms. Elsewhere (Bach 2011), I have suggested that the scientific-theory-theory account can deliver an adequate response to nativist challenges if it is supplemented by a structure-mapping account (see Gentner 1983; Gentner & Colhoun 2010) of hypothesis discovery.
References


