The Philosophy of Social Cognition

An Introduction to the Philosophical Approaches of Mentality and Sociality


Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée

www.hardyvallee.net/POSC
# Table of Content

Table of Content ................................................................................................. 3  
Introduction ........................................................................................................ 5  

**PART ONE: PHILOSOPHY AND THE OTHER MINDS** ........................................ 6  
1 The Other Minds................................................................................................ 7  
1.1 The Problem of Other Minds ........................................................................ 8  
1.2 The Egocentric Conception .......................................................................... 9  
1.3 The Detached Conception ............................................................................ 10  
1.4 Comparisons .................................................................................................. 12  
1.5 References ...................................................................................................... 14  

2 Rationality and Interpretation .......................................................................... 15  
2.1 The Varieties of Rationality .......................................................................... 16  
2.2 The constitutive ideal of rationality ................................................................. 18  
2.3 The intentional stance .................................................................................... 21  
2.4 Ideals and Stances .......................................................................................... 22  
2.5 References ...................................................................................................... 24  

**PART TWO: OUR EVERYDAY PSYCHOLOGY** .................................................... 26  
3 Simulation and Theory-Theory ......................................................................... 27  
3.1 Some Conceptual Clarifications ..................................................................... 28  
3.2 Theoretical minds .......................................................................................... 30  
3.3 From External to Internal Theory ................................................................... 30  
3.4 The Theory-Theory: ....................................................................................... 32  
3.5 Simulating minds ............................................................................................ 33  
3.6 A Posteriori Simulation .................................................................................. 34  
3.7 A Priori Simulation ........................................................................................ 35  
3.8 Comparisons ................................................................................................... 36  
3.9 References ....................................................................................................... 38  

4 Embodied Phenomenology and Narratives ....................................................... 40
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 The Critique of Folk-Psychological Reason</td>
<td>41</td>
</tr>
<tr>
<td>4.2 Phenomoeology and the limits of “theories of mind”</td>
<td>42</td>
</tr>
<tr>
<td>4.3 Folk-psychological narratives</td>
<td>44</td>
</tr>
<tr>
<td>4.4 References</td>
<td>46</td>
</tr>
<tr>
<td>5 Experimental Approaches to Folk-Psychology: Moral Judgments and Pluralistic Accounts</td>
<td>47</td>
</tr>
<tr>
<td>5.1 Folk-Psychology, From Intuitions to Experimentation</td>
<td>48</td>
</tr>
<tr>
<td>5.2 Pluralism in Folk-Psychology</td>
<td>49</td>
</tr>
<tr>
<td>5.3 The Experimental Philosophy of Folk-Psychological Judgment</td>
<td>50</td>
</tr>
<tr>
<td>5.4 Folk-psychological accounts : a synthesis</td>
<td>54</td>
</tr>
<tr>
<td>5.5 References</td>
<td>54</td>
</tr>
<tr>
<td>PART THREE: THE CONSTITUTION OF THE SOCIAL MIND</td>
<td>56</td>
</tr>
<tr>
<td>6 Neurons that Mirror</td>
<td>57</td>
</tr>
<tr>
<td>6.1 Explaining the social mind</td>
<td>58</td>
</tr>
<tr>
<td>6.2 The brain</td>
<td>59</td>
</tr>
<tr>
<td>6.3 The Mirror Neurons System (MNS)</td>
<td>61</td>
</tr>
<tr>
<td>6.4 Mirroring and interpreting</td>
<td>62</td>
</tr>
<tr>
<td>6.5 References</td>
<td>63</td>
</tr>
<tr>
<td>7 Social Primates and the Evolution of Social Cognition</td>
<td>65</td>
</tr>
<tr>
<td>7.1 Evolution and social cognition</td>
<td>66</td>
</tr>
<tr>
<td>7.2 Primate minds</td>
<td>67</td>
</tr>
<tr>
<td>8 The Modularity of The Social Mind</td>
<td>72</td>
</tr>
<tr>
<td>8.1 Modularity and the Mind</td>
<td>73</td>
</tr>
<tr>
<td>8.2 Folk-psychology and modularity</td>
<td>74</td>
</tr>
<tr>
<td>8.3 Evolutionary psychology and modularity</td>
<td>75</td>
</tr>
<tr>
<td>8.4 The limit of the Modularity Thesis</td>
<td>77</td>
</tr>
<tr>
<td>9 Social Intelligence</td>
<td>79</td>
</tr>
<tr>
<td>9.1 Human Intelligence: puzzle, chess or collective hunting?</td>
<td>80</td>
</tr>
<tr>
<td>9.2 Ecological and Social Intelligence</td>
<td>80</td>
</tr>
<tr>
<td>9.3 Socio-ecological Intelligence</td>
<td>82</td>
</tr>
<tr>
<td>Bibliography on Philosophy and Social Cognition</td>
<td>86</td>
</tr>
</tbody>
</table>
Introduction
PART ONE: PHILOSOPHY AND THE OTHER MINDS

As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation.

- Adam Smith, The Theory of Moral Sentiments
1 The Other Minds


Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
1.1 The Problem of Other Minds

As Aristotle pointed out a long time ago, we are by nature political animals, i.e., animals that live together, such as the bee, the wasp or the crane. If it’s in our nature, then we must expect our minds to exhibit some proficiency in social cognition, such as social perception, social reasoning, social learning, communication, knowing what others think, managing complex information structures related to social phenomena such as “Phoebe thinks that Monica knows that Phoebe knows that Monica and Chandler are together”. These processes and cognitions presuppose that we know that others have a mind. But how do we know (if knowledge is justified true belief) that others have a mind and what they have in mind?

It could be an inference to the best explanation: our best theory of mind (either intuitive or scientific) may imply that there are no reasons to doubt that other things that look like minded human beings are in fact, minded human beings. Similarly, it could be an analogical inference. Both are interwoven here in David Hume’s argument to attribute reasons to animals:

> We are conscious, that we ourselves, in adapting means to ends, are guided by reason and design (...) When therefore we see other creatures, in millions of instances, perform like actions, and direct them to the ends, all our principles of reason and probability carry us with an invincible force to believe the existence of a like cause. (A Treatise of Human Nature, Sect. xvi.)

Or it could be a conceptual link. It is analytical of, or deeply entrenched in the grammar of, or a rule in the language-game of mental state attribution, that the link between behavior and mental states is not inductive nor an entailment, it is conceptual.

In any case, the problem is interesting for many disciplines. Philosophy of mind wonders how social understanding is possible; epistemology ask how this knowledge could be justified; social science and the philosophy of social science question whether the knowledge of other is something specific to human sciences or if it’s in continuity with other sciences; cognitive science investigate the mechanisms by which we come to this knowledge. In any domain, the central controversy, as Stueber puts it, is

> whether our epistemic access to other minds proceeds in a manner radically different from that by which we acquire knowledge about other domains of investigation, or whether it is structurally similar to such methods (Stueber, p.2)

They are two generic answers to this question because there are two generic ways of understanding other minds. Stueber gives the following example: how can we know whether two people are the same height? We can use two types of standards: an external standard such as a measuring tape or an internal standard, i.e., using yourself as a standard when you stand back to back with someone else. Similarly, in interpreting a text, a sentence or an action, we may use internal or external standards. In the first case, we put ourselves in the actor’s shoes, and use our “capacity to transform ourselves imaginatively into an occupant of the other person’s point of view” (Stueber); we change place in fancy as Adam Smith formulated it (Theory of Moral Sentiments, I, I, i). I will call it (following Stueber) the engaged conception. In the second case (the detached conception), we do not use our ability to reconstruct the target’s (the agent we
interpret) point of view, but recruit some knowledge and apply it to the situation. The first one thus suggests that there is something peculiar to social understanding, something that is not found in other domains.

1.2 The Egocentric Conception

The first conception could be called egocentric, empathic, first-person, subjective, engaged, or understanding-based. It has roots in the romanticist view of the world and foundational debates in hermeneutics and social science. Until the 17th and 18th century, hermeneutics (the theory of interpretation) was a logic of interpretation and argumentation, a set of tools for reasoning that complemented basic logical skills. In the 19th century, it shifted focus from the logic of interpretation to the process of interpretation (e.g. how do we locate a text in an historical narrative, a speech in a religious ceremony, etc.) and the justification of interpretation. These questions led Dilthey to induce a psychological turn in the theory of interpretation: in philology and all the human/social sciences, we must use a particular method that is not found in other domains of enquiry: empathy. Contrary to Newton or other philosophers who wanted to introduce Newtonism (i.e. mechanistic and nomological explanations) in the social domains, Dilthey argued that human matters are to be understood, not explained: when we want to understand why Napoleon wanted to invade Russia or what Shamanic beliefs represent for a certain Native American cultures, we cannot evoke laws of nature or mechanisms (such as associationist psychological mechanisms), but reasons. Empathy-based interpretation, or re-enactement, is at the root of all human/social science. Whereas natural science has a nomological mode of explanation (a subsumption under a law), social science tries to reconstruct the holistic structures of the reasons; understanding a sentence, a text, or an action requires the “big picture” (context, society, history, etc.). We understand Napoleon’s decision and a Native American belief by showing the network of beliefs, desires and cultural factors that gives meaning to a particular action or sentence.

The engaged conception thus emphasizes the peculiarity of understanding other minds; we see others as minded and as same-minded, we use our knowledge of ourselves as minded agents to model someone else's mind, a peculiar form of social reasoning, different to the reasoning employed when we predict/explain physical phenomena. This conception has a normative and a descriptive dimension. The normative dimension is a suggestion that empathy is an epistemically central process for understanding other minds; i.e., that empathy can give us knowledge, justified true beliefs, about what others thinks. On its descriptive side, the engaged conception construes empathy as an inner imitation for understanding other minds: interpreting requires that the mental process by which one produces an action/utterance are the same process that are recruited by the comprehension of that action/utterance. As Ryle argued, “the knowledge that is required for understanding intelligent performances of a specific kind is some degree of competence in performances of that kind” (1949, 51). Understanding is a skill rather than a theory, knowing-how rather than knowing-that.

Empathy can have an epistemic function (knowing what others think, feel, etc.), an evaluative function (approving or disapproving others’ reactions) and a rational function: the development of self-criticism by recognizing the opinions of others about myself.

One can distinguish basic and reenactive empathy. Basic empathy is the quasi-perceptual inference to the attribution of mental states, such as when we know that someone is angry; the
other one is a deliberative reconstruction of reasons, such as when we know why she is angry. Empathy is thus not just a fast, automatic and intuitive process, but also a more complex form of reasoning by which we reconstruct someone else’s perspective and her network of reasons, beliefs and desires.

Sympathy is a related construct. While empathy is primarily an epistemic notion, sympathy is an epistemic and moral notion. Since Adam Smith and David Hume, it refers to the connection between acting morally and the intuition (or perception, or imagination) of the pain and pleasure of others. It could be described as a concern about other minds, rather than the knowledge of other minds. Sympathy, or “fellow-feelings” has a central role in social understanding: “we can form no idea of the manner in which [other persons] are affected, but by conceiving what we ourselves should feel in the like situation.” (TMS, I, I, i). It is also a normative notion, because in sympathy, we also approve or disapprove of others’ reactions; we also assess each other’s intellectual and affective judgment according to our feelings. It is thus a convenient judgmental tool, the only one we have, according to Smith:

*I judge of your sight by my sight, of your ear by my ear, of your reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them.* (I, I iii)

The main problem that face the engage conceptions are the following:

1. How to justify empathic information? is it knowledge?
2. How to reconcile it with complex talks like a discussions about two characters?
3. What if we don't always put ourselves in others’ shoes?

### 1.3 The Detached Conception

The second conception could be called non-egocentric (allocentric), theoretical, third-person, objective, detached, or explanation-based. Its principal feature is to construe our knowledge of other mind not as some sort of simulation skill or re-enactive process, but as the application of an intuitive, commonsense psychological theory.

Sellars ask us to imagine fictional humans speaking a language limited to the “public properties of public objects” (XII). They can only point what is perceptible to everybody. They can tell you have a red shirt, but not that they “believe” that you have a red shirt, or that they “want” to have a shirt like yours. What would this language require in order for its locutors to be able to express “inner”, unobservable mental episodes? In order to be able to talk about mental states, this language and these individuals need an intuitive theory. Theories are not confined to scientific research. If a theory is a body of structured knowledge about phenomena and their cause, then we explain empirical phenomena with similar epistemic resources. As Sellars puts it:

*science is continuous with common sense, and the ways in which the scientist seeks to explain empirical phenomena are refinements of the ways in which plain men, however crudely and schematically, have attempted to understand their environment and their fellow men since the dawn of intelligence.* (Sellars, XIII)

The detached conception sees the interpretation of action as an explanatory process, similar to the explanation of other phenomena. While the engaged conception see interpretation
as having a peculiar operational mode, empathy, the detached conception see it as an instance of 
a more general capacity, where theoretical knowledge and information about a context are 
integrated so as to produce an explanation. A theory requires certain cognitive resources not 
ecessarily involved in empathy: a certain representational format (structured network of 
propositions) and a certain inferential mode (inferences aimed at explaining a phenomena). 
Theories seem to require a faculty of judgment distinct from sensation, perception, emotions or 
imagination (all aspects of general perceptive capabilities): contrary to Adam Smith who saw 
sympathy as the only mean by which we can understand and assess each other, the detached 
conception argues instead that perception or imagination is unable to represent our knowledge of 
other minds. As Descartes argued, our knowledge of things does not come from perception or 
imagination, but rather from judgment. I can imagine a \textit{pentagon}, but not a \textit{chiliogon} (1000 
 sides): the difference between them comes from my propositional knowledge that a pentagon has 
5 sides and that a chiliogon has 1000 sides. This reasoning applies to other humans as well:

\begin{quote}
what do I see from the window beyond hats and cloaks that might cover artificial 
machines, whose motions might be determined by springs? But I judge that there are 
human beings from these appearances, and thus I comprehend, by the faculty of judgment 
alone that is in the mind, what I believed I saw with my eyes (Second Meditation)
\end{quote}

This theory, or folk-psychology, is a set of common sense “platitudes”: obvious, shared, 
intuitive generalizations which are common knowledge amongst us, such as: that people do not 
like pain, that folk believe and desire, etc. This intuitive theory is similar to folk physics or folk 
biology: it is a non-scientific version. It is a “theory of mind implicit in our everyday talk about 
mental states (...) remarks linking sensory experiences to mental states” (Ravenscroft, 2004; see 
also Lewis 1972)

We’ve seen that the engaged conception can be a theory of mental process (we interpret 
by reconstructing) and a theory of social-scientific inquiry (human beings are not studied like the 
rest of the natural world). Similarly, the detached conception can be construed as such. As a 
three of mental process, it states that interpretation is the application of and intuitive 
psychological theory that does not require an egocentric perspective. As a theory of social-
scientific enqiry, the detached conception is close to naturalism: it does not see social sciences 
as a separated domain of enquiry that requires a particular methodology such as re-enactment, but 
stresses instead the continuity of social science with natural science. From Hume to the Logical 
Positivists, proponents of the detached conception tried to bring natural-scientific methodology in 
social science. Newton established a new standard of rigor and simplicity in explanation, and 
many attempts to naturalize social science can be understood as attempts to bring Newtonian 
methodology (finding simple laws expressed that explain and predict a lot of phenomena) in this 
field. Empathic understanding, for proponents of the continuity between natural and social 
science, is not a particular methodology specific to the former, but a convenient tool, a useful 
heuristic whose utility lies in its ability to lead to general scientific laws about human behavior 
(see Hempel & Oppenheim, 1948). As Neurath (one of the member of the influent Circle of 
Vienna) puts it:

\begin{quote}
\textit{Empathy, understanding (Verstehen) and the like may help the research worker, but they 
enter the totality of scientific statements as little as does a good cup of coffee which also 
furthers a scholar in his work}’ (Neurath, 1931, p. 357).
\end{quote}
The detached conception face the following problems

Who says belief-desires are central?

Who says we use theory?

Who says this theory is descriptive/predictive and not normative?

What about character traits attributions? conscious states, etc?

Do we really have a theory that says that *if someone cries, then she must be sad* and apply this theory when we see people crying, or is there a more basic mode of perception involved here?

### 1.4 Comparisons

The two conceptions presented here (the engaged and the detached) are in fact two generic families—not always exclusive—of theory of interpretation and social cognition. They were characterized broadly because each individual philosopher might not fit squarely in one family. One could, however, identify certain features that are typically associated with each family. The following table illustrates what proponents of each conception tend to adopt, regarding the proper conduct of social-scientific investigation and the nature of the mental process of interpretation. Each of these topics can be divided in sub-topics:

**Philosophy of Social Science**

*Epistemic goal*

What is the goal of the study of human beings and their intentional productions? To reconstruct their meaning, to locate them in a web of belief (understanding) or is it the same epistemic goal that all science shares? (explanation)

*Unit of analysis*

What should figure in interpretation: reasons or causes? Mental contents and representations or causal entities?

*Nature of social/human science*

Is social science continuous with or distinct from natural science?

*Mode of enquiry*

In interpreting text and people, should we re-enact their reasons, or subsume them under general theories?

**Mental Processes**

*Perspective*

Do we interpret from a first-person, egocentric or a third-person, allocentric perspective?

*Cognitive mechanism*

Do we use a simulation skill or an intuitive theory?

*Scope of the mechanism*
Does the mechanism process only social information (domain-specific) or is it a mechanism that can be activated in other domains? (domain-general)

The following table illustrates the typical views of each conception:

<table>
<thead>
<tr>
<th>Features</th>
<th>Engaged Conception</th>
<th>Detached Conception</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Philosophy of Social/Human Science</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epistemic mode</td>
<td>Understanding</td>
<td>Explanation</td>
</tr>
<tr>
<td>Unit of analysis</td>
<td>Reasons</td>
<td>Reasons and Causes</td>
</tr>
<tr>
<td>Nature of social/human science regarding the other science</td>
<td>Exceptionalism</td>
<td>Continuism</td>
</tr>
<tr>
<td>Mode of enquiry</td>
<td>Re-enactment</td>
<td>Theoretical reasoning</td>
</tr>
<tr>
<td><strong>Mental Processes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perspective</td>
<td>Egocentric Subjective</td>
<td>Allocentric Objective</td>
</tr>
<tr>
<td>Cognitive mechanism</td>
<td>Simulation skill <em>(knowing-how)</em></td>
<td>Intuitive theory <em>(knowing-that)</em></td>
</tr>
<tr>
<td>Scope of the mechanism (bold indicate a preference)</td>
<td>Domain-specific/general</td>
<td>Domain-specific/general</td>
</tr>
</tbody>
</table>
1.5 References

http://en.wikisource.org/wiki/Meditations_on_First_Philosophy


Smith, A. ([1759] 2002). *The Theory of Moral Sentiments*
http://www.adamsmith.org smith/tms/tms-index.htm (Adam Smith Institute)

http://mitpress.mit.edu/books/chapters/026219550Xintro1.pdf
2  Rationality and Interpretation
2.1 The Varieties of Rationality

_all of philosophy (...) is almost coextensive with the theory of rationality_  
(Putnam 1981, pp.104-105)

The detached and the engaged conception of interpretation recommend two different approaches: one based on empathic understanding, the other on theoretical explanation. Although these two conceptions frame almost all the debates regarding the nature of interpretation, another concern raises many theoretical problems: does interpretation (whether as understanding or as explanation) presuppose the rationality of the interpreted agent? And if so, to which extent? These are the issues tackled in this chapter.

Rationality is a complex, if not confused, concept. It is thus important to explore first different meanings of rational. First, one can distinguish rational agents and rational actions/thoughts. Rational agents can produce rational or irrational actions/thoughts, while non-rational or a-rational entities just cannot produce any kind of actions/thoughts. Thus rational and irrational actions/thoughts are produced by rational agents.

The actions and thoughts of rational agents are rationally assessable: they can be said to be correct or incorrect, optimal or not, justified or not, etc. Rationality is normative, and this evaluation can target either thoughts or actions. The first type is theoretical rationality: the rationality of inferences, thoughts, reasoning, beliefs (formation, revision, justification), mental representations, theories and explanations. The second one is practical rationality: the rationality of actions, behaviors, decisions, intentions, motivations, preferences and strategies. In sum, what to think vs. what to do. This distinction requires two sets of norms: logic, epistemology, semantics, probability theory for the first, decision theory, game theory, general equilibrium (a.k.a. market) theory for the second. Both types of rational evaluations also imply different forms of irrationality, for instance regarding arbitrary choices and wishful thinking. Regarding the former, it is theoretically irrational to believe A and not B when A and B are equiprobable, although it can be practically rational to prefer A to B when A and B are equally satisfying (not choosing can be practically irrational). Regarding the latter, it is theoretically irrational to let beliefs be influenced by desires, but it can be practically rational to let actions be influenced by desires (Harman, 1999).

Rational evaluation of practical rationality can be internal or external. In the first case, we assess the coherence of intentions, actions and plans. Actions make sense “from the point of the cognitive and conative perspective of the agent” (Stueber 2006, p. 49). In the second case, we assess the effectiveness of a rule or procedure for achieving a certain goal. Actions make sense “relative to a given set of environmental parameter that include the agent's desire but not his belief “ (Bermúdez, 2002, p. 260). An action can be rational from the first perspective but not from the second one, and vice-versa. Poor performance in probabilistic reasoning can be internally rational (subjects may have good reason to choose a certain prospect) without being externally rational (their behavior is still suboptimal). The Gambler’s fallacy is and will always be a fallacy: it is possible, however, that fallacious reasoners follow rational rules, maximizing an unorthodox utility function. There are thus two ways—non-exclusive—to behave irrationally. One can be externally irrational if the outcome of an action is considered suboptimal; in this case
the attribution of irrationality requires data about the agent and the outcome. (ex: Gambler's fallacy). One can also be internally irrational if, regardless of the outcomes of the action, the agent’s desires and action performed are incoherent. Akrasia (acting against one’s best judgment) is a form of practical irrationality because a desire (I want to stop smoking) and the action (lighting up a cigarette) are not coherent.

Many philosophers suggested that, in order to interpret another agent, a sentence or an action, we implicitly presuppose the **rationality** of the agent. This **assumption** modulates our interpretation and makes it possible. As Sorensen summarize the idea, “[a]nyone who superimposes the longitudes of desire and the latitudes of belief is already attributing rationality” (Sorensen, 2004p. 291). The claim can have a **weak** reading: agents are “not stupid”: they have reasons to act when they behave intentionally. On a **stronger** reading, it suggest that agents should comply with rationality standards. The latter is common in classical economics: it is assumed that agents are utility-maximizer (just like physics assumes frictionless bodies or ideal gases). It can also be an **interpretative** postulate: rational agents are rationally interpretable, and rationality is a justified or useful supposition of the coherence between beliefs, desires and actions. The assumption is thought variously to be a possibility condition, an empirical hypothesis or a superfluous statement. For Davidson, it is necessary for the **application** of concepts such as action, belief, desire, intention; For Dray, it is necessary for **interpreting** action, belief, desire, intention. For Hempel (1962, p.12), it is an empirical assumptions that figures in an explanation:

A was in a situation of type C
A was a rational agent
In a situation of type C any rational agent will do x

Therefore A did x

Two important theoreticians of rational interpretation, Donald **Davidson** and Daniel **Dennett** (presented in detail in the forthcoming sections), articulated and refined the idea that rationality is also an interpretive norm (one might say a meta-norm), more fundamental than particular axioms or formal systems. For Davidson and Dennett, rationality is not a precise set of rules to follow, but a condition for being evaluated according to such rules. To be a rational agent does not imply a perfect rationality, but being assessable with rationality standards; as Stueber puts it, it is to be “normatively required to be responsive to norms of reasoning” (Stueber 2006, p. 63). In the practical domain, it supposes, among other things, that we consider the agents’ beliefs and desires as reasons for action; in the theoretical domain, it supposes, among other things, that we consider the agents’ beliefs and desires to be coherent.

The rationality postulate is compatible with the engaged and detached conceptions. One can argue that empathy always simulates rational agents; or that one of the assumptions of a theory of mind is the rationality of the agent. Some accounts concern the mental process of interpretation (e.g. Dennett); in this case, the rationality assumption implies that we have an intuitive concept of “rational agent” and that it is involved in interpretation. Other accounts (e.g. Dray) are about the method of the social scientist: she must presuppose the rationality of the target agent.
2.2 The constitutive ideal of rationality

Davidson ideas are directed at logical positivism and Wittgensteinian philosophy. Against positivism, Davidson holds that intentional actions cannot be explained like other empirical phenomena because they involve reasons. Against Wittgenstein, he holds that intentional actions are also causal relations (but cannot be explained under the Deductive-Nomological model), mental objects are material objects (but cannot be described by scientific laws, a view named the “anomalism of the mental”), and that reasons are cause. Actions are particular events (such as in: the rock broke the window; no law is needed to explain it), caused by mental causes (beliefs, desires); these causes are also reasons that justify (or imply) the action when the action is rational. An irrational action is an action whose reasons causes the action but do not justify the action.

According to the Deductive-Nomological (DN) model, one explains an empirical phenomenon by showing how it “fit into a nomic nexus” (Hempel 1965:488). If, from the statement of a scientific law or a universal generalization (called also a covering law) and certain conditions one can logically deduce an event, this event is then regarded as being explained:

(1) L (scientific laws or universal generalization)
(2) C (condition)

---------------------------------------------------
(3) E (event)

If rationality was an empirical assumption, then action explanation would be of this kind:

If something is soluble it dissolves in liquids of a certain sort
Warm coffee is such a liquid,
All sugar is soluble,
That this cube is sugar

-------------------------------------

A particular small cube dissolved in warm coffee

This pattern of explanation requires general knowledge about sugar.

But if rationality is constitutive, then action explanations are particular explanation, grounded in local facts:

This cube was soluble
Soluble things dissolve in coffee
This cube was in coffee.

---------------------------------------------------

This particular small cube dissolved in coffee

Thus, to explain someone's action, we don't need (or cannot have) laws of behavior, but norms of interpretation, just like we don’t need much general knowledge in the last deduction.
According to Davidson, the mental and the physical have different **constitutive** concepts: the mental in grounded in rationality. Rationality is a condition of interpretation, it is almost like a Kantian category of the understanding: it is not derived from experience, but rather makes it possible. For instance, if subjects in an experiment prefer a 100% to win 100$ than 50% to win 300$, “by the book”, subjects are irrational: they prefer the option with the lowest utility. But again, certainty can have a value, therefore it is not irrational. Since it is so easy to rationalize, rationality is not an empirical concept:

(...) the satisfaction of conditions of consistency and rational coherence may be viewed as constitutive of the range of application of such concepts as those of belief, desire, intention and action

(...) if we are intelligibly to attribute attitudes and beliefs, or usefully to describe motions as behaviour, then we are committed to finding, in the pattern of behaviour, belief and desire, a large degree of rationality and consistency

*(Essays on Actions and Events, p.237)*

An important point in Davidson theory, is that rationality and rational interpretation requires linguistic communication. His argument goes as the following:

Argument:

(1) To be a rational animal is to have propositional attitudes

(2) Propositional attitudes are organized logically and systematically.

(3) Since we consider our propositional attitudes as being true or false, correct or incorrect, we have the concepts of belief, meaning and truth

(4) In order to rationalize action (as in 1), attribute propositional attitudes (as in 2) and possess the concepts mentioned in (3) a creature must be linguistic

(1) *To be a rational animal is to have propositional attitudes*

The argument for Davidson (1) is that when we describe the behavior of a person in terms of beliefs and desires, we show the rationality of the action in the light of the content of belief and the object of desire. The structure of the beliefs, desires and actions of Alice, in connection with her desire for apples, shows that Alice is rational because it is a coherent structure of reasons. It is therefore impossible to attribute beliefs and desires in the first place, and then, after the fact, to add that this creature is rational and to consider the last statement as informative. One cannot say that (1) “Mike took a taxi because he wanted to be in time and knew that the bus would take too much time” and saying after that (2) “Mike is rational” is informative: (2) is implicit in (1). If he has beliefs and desires, he is rational.

For example, if someone pulls a rope by its two ends, an explanation of this action that does not presuppose the rationality of the agent could be that she is fighting against herself. There are many ways to describe it as a rational action but we need to add several auxiliary hypotheses, whereas if we assume the consistency of desires, beliefs and actions, the assumption that she tries to break the rope is more natural. The attribution of propositional attitudes is thus made possible by the **constitutive ideal of rationality**

(2) *Propositional attitudes are organized logically and systematically.*
If one accepts (1), one has to accept (2): to have one propositional attitude is to have a panoply. The content of a belief is the sum of its relationship with all other beliefs. In order to individuate beliefs, the interpreter must locate them in a network of reasons. If an agent has the concept CAR, and that this concept specifies that a car is a vehicle, then it should be in principle possible to attribute to an agent who believes that she is near a car, the belief that she is near a vehicle. Otherwise, she just does not have the concept of CAR. If one has the concept of car, one has many general beliefs: they usually have four wheels, they go on the road, they are powered by an internal combustion engine, they are able to carry a small number of people, etc. Whatever is part of the possession of the concept CAR must figure in the set of beliefs we can attribute to this agent, otherwise it makes no sense to attribute her the concept of CAR, and we cannot make the difference between “she believes that there is a car” and “she believes that there is a big metallic object”, and we cannot attribute her false beliefs, such as “she believes wrongly that this is a car while in fact it is a sculpture”. Intentional states are typically intentional: from (1) Ana believes that Paris is in France, and (2) Paris is the capital of France, it does not follow that (3) Ana believes that Paris is the capital of France.

(3) Since we consider our propositional attitudes as being true or false, we have the concepts of belief and truth

The concept of BELIEF is a necessary condition for the possession of beliefs: to believe that P implies also to believe that you believe that P. It is believed that it is believed that P when there is a possibility of surprise. If I am surprised that no coin was in my pocket, the surprise is impossible if, in the first place, I did not have a belief that there was a coin in my pocket. Thus when I am surprised, I now believe that my former belief was wrong, which supposes that I believed that I believed that I had a coin in my pocket. Surprise implies a contrast between what is considered to be true, and what is actually true, i.e. a distinction between truth and falsehood.

(4) In order to rationalize action (as in 1), attribute propositional attitudes (as in 2) and possess the concepts mentioned in (3) a creature must be linguistic

Davidson argues that to be interpreted rationally, a target agent must act on the basis of propositional attitudes; however, the possession of beliefs presupposes several other beliefs, according to (3), and even the concept of BELIEF. This concept presupposes the language capability. Indeed, if one defines belief as a thought that is likely to be semantically evaluated (it can be true or false), creatures that entertain beliefs have the idea of an objective independent reality and are able to distinguish the objective and the subjective. But this distinction is guaranteed, according to Davidson, by the mastery of language. While a non-linguistic creature can at best learn and generalize, nothing in its behavior is a basis for concluding that it can distinguish the objective and the subjective. One might conclude that this is a response to different classes of stimuli, but no more. Linguistic creatures (humans) have to communicate, to be able to think that the other communicator also has the concept of an outside objective world.

Possession of a language would constitute a condition for the use of the concept of rationality, according to Davidson: “rationality is a social trait. Only communicators have it” (1982: 327). Being a linguistic agent allows us to use prior and passing theories about other speakers, and the use and revision of these theories are governed at each step by the constitutive ideal of rationality:
For the hearer, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker, while the passing theory is how he does interpret the utterance. For the speaker, the prior theory is what he believes the interpreter's prior theory to be, while his passing theory is the theory he intends the interpreter to use. (1986, 442)

Agents have theories about what other agents mean, and revise these theories in the course of interaction. Thus we interpret by employing our constitutive ideal of rationality. The norms of rationality are norms of interpretation and also norms that govern the use of intentional concepts.

2.3 The intentional stance

While Davidson was preoccupied by the pervasiveness of rationality in interpretation because of the open-ended possibility of justifying an action (as the practice of experimental economics show), Dennett was more preoccupied by the pervasiveness of rational interpretation displayed by cognitive science: computers, babies, animals and linguistic beings can all be described as information processors. As he notes, the notion of possession of “information or misinformation is just as Intentional a notion as that of belief” (Dennett 1971, p.90). They all can be described as wanting and desiring, but when is it correct? While Davidson was more interested by the justification of interpretation, Dennett is more interested by the utility of interpretation.

The concept of rationality, in Dennett's view, is articulated in the context of predictive strategies, intuitive attitudes or stances with respect to certain types of system with which we can interact (similar to Newell's (1990) principle of rationality). Whether it's an apple falling from a tree, a software failure or a seller of vacuum cleaners, as soon as we come into contact with these objects or agents we have a tendency to generate certain predictions about the evolution of the target object or agent (the “system”):

1. The apple, if there is no wind, will fall in a straight line toward the ground
2. The chess software will take our pawn
3. The vacuum cleaner seller will recommend that we buy one.

We have prior and passing theories for many kinds of systems. We carry out different kinds of predictions as the system is predicted and explained from the physical stance, the design stance, and the intentional stance. In the first case (an object which falls), our presuppositions are limited to the physical features of the objects. In the second case, our presuppositions are limited to functions (the algorithms of a computer, or the biological functions of the organs). In the last case, we presuppose the rationality of the agent, namely that the system in question is equipped with beliefs and desires:

“Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent will do.” (Daniel Dennett, 1987, p. 17)
“The success of the stance is of course a matter settled pragmatically, without reference to whether the object really has beliefs, intentions, and so forth; so whether or not any computer can be conscious, or have thoughts or desires, some computers undeniably are intentional systems, for they are systems whose behavior can be predicted, and most efficiently predicted, by adopting the intentional stance toward them.” (Dennett 1978, p. 238)

The strategy can be employed with many different systems: computers, babies, animals, humans, etc. What matters is the success of the stance. Dennett gives the example of the plover and its “deceptive” behavior when it detects a predator; it behaves as if its wing was broken, and “pretends” to be injured. Is the deception a real intentional deception, which reveals the rationality of the agent, or a simple reflex? To examine this (Dennett, 1987: 258) proposes to interpret the behavior of the plover as a soliloquy, where the animal intentionally tries to fool the predators. If we can successfully use the intentional stance, if it shows that the animal is sensitive to a complex set of environmental conditions that the soliloquy suggests, we can say that it entertains beliefs and desires. If these tests show that cognitive control mechanisms react to simple environmental patterns, then it is necessary to reassess its rationality downwards. Computers are another interesting case: we say that the computer wants to take my pawn, but this ascription is useful or not useful, not right or wrong: it is just more convenient to say that than to describe all its algorithms and internal states. This procedure is the same, according to Dennett, as we use between ourselves: we assume that agents are informed and capable of logical inferences, predict their behavior, and reassess their rationality if necessary.

Stances are not appropriate and relevant for any kind of system:
the physical stance is not useful when we want to talk about the functioning of a software
the design stance is not useful when we talk about the fall of an object
the intentional stance is not useful when we talk about the movement of clouds, or the trajectory of a missile

It is useful when it picks up regular patterns. On the other hand, we can adopt “the Intentional stance in one's role as opponent, the design stance in one's role as redesigner, and the physical stance in one's role as repairman” (1971, p. 91).

Thus the intentional stance is both a cognitive mechanism and a social-scientific methodology. It is grounded in natural selection: if evolution has done its job, our prediction will be useful (1971, p.93). Our folk-psychology is thus an adaptation. Social science must proceed by assuming adapted design, and going from the intentional to the design stance. Instead of restricting intentionality to language, Dennett restricts it to (useful attribution of) information-processing. One can call it evolutionary rationalism: natural selection designed us to “get the answer right and to want to get it right” (97). Thus beliefs, desires and so on are like centers of gravity: not “Real”, but useful, like gravity centers in physics.

2.4 Ideals and Stances

Both Dennett and Davidson are continuers of Quine and the pragmatist tradition, but with different inspirations and aspirations. Davidson faced the problems of the irreducibility of rationalization in experimental economics; Dennett saw how the intentional stance works in so
many circumstances with humans and non-humans. Davidson focuses on interaction with other humans, Dennett with other “systems”; Davidson on justification, Dennett on usefulness. They have different conceptual grounds: for Davidson it is logic and semantics, for Dennett, evolutionary biology and cognitive science (where non-linguistic agents are intentional agents). They also have different explanatory “target”: linguistic communicator vs. information-processing systems. They also had different impact: Davidson on social science and literary studies, Dennett on cognitive science and biology.

However, both take rationality as a fundamental, sui generis notion (linked with beliefs, desires, prediction and explanation) that cannot be completely codified: it is an ideal notion, but this ideal is not the economist's rational agent. It is more like the warrant of commonsense psychology. It is “a good sense of when to rely on what” (Dennett, 1987, p.97). Beliefs and desires are “abstracta”. There are no basic principles of rationality (Davidson, 2004 :196). Both see beliefs and desire ascriptions as interlocked, and attributing rationality as a faster way of interpreting. For Davidson, it is faster than irrational interpretation (e.g. see the rope example). It is easier to see it as a rational action than a non-rational one. For Dennett, it is faster than physical stance interpretation (e.g. predicting a computer's move based on quantum physics). Both see irrationality as meaningful only in the light of rationality. For Davidson, one has to suppose some partition in the mind (“semi-independent structures” see (Davidson, 1982b)). For Dennett, we cannot attribute irrationality from the intentional stance, but need the intentional stance to understand malfunctioning psychological mechanisms and their role in our holistic cognitive economy. Both see questions of interpretation as ubiquitous and as an everyday competence based on an intuitive theory.

One of their main differences is about mental and non-mental property. Davidson opposes the mental and the physical, as two domains based on different constitutive concepts. Dennett see that there are 3, not 2, domains (the physical, the functional and the intentional), and that although they have their own properties, one can use a higher-level domain to investigate the lower-level. Davidson sees intentionality, rationality and normativity as tied together. No language, no mind. Dennett advocates a continuist position. They also have different stances on psychological explanation: for Dennett we can go on and mechanically decompose intentional systems in subsystems. For Davidson the holism, the anomalism of the mental and its irreducibility to the physical makes that impossible. This difference is probably because Dennett links intentionality (and rationality) to information rather than language and semantics. Both see the problems of over-rationalization (attributing rationality to non-rational system such as missiles or flowers). For Davidson, it is unjustified, for Dennett, it is not useful.

Both are more on the detached side. For Davidson, interpreters adjust their prior/passing theories and the mental can only be understood through particularist explanation (where no generalizations are involved), as in “the window broke because it was struck by a rock” (Davidson, 1963). We rationalize actions with beliefs and desires under the ideal of rationality, but not causes. For Dennett, when we interpret, we have a theory of behavior (1971, 93):

“An interesting idea lurking in Stich's view is that when we interpret others we do so not so much by theorizing about them as by using ourselves as analog computers that produce a result. Wanting to know more about your frame of mind, I somehow put myself in it, or as close to being in it as I can muster, and see what I thereupon think (want, do. . ). There is much that is puzzling about such an idea. How can it work without being a kind of theorizing in the end? For the state I put myself in is not belief but make-believe
belief. If I make believe I am a suspension bridge and wonder what I will do when the wind blows, what “comes to me” in my make-believe state depends on how sophisticated my knowledge is of the physics and engineering of suspension bridges. Why should my making believe I have your beliefs be any different? “(1987, P. 100)

2.5 References


PART TWO: OUR EVERYDAY PSYCHOLOGY

Our ancestors begin to speak of the privileged access each of us has to his own thoughts. What began as a language with a purely theoretical use has gained a reporting role.

- Wilfrid Sellars, Empiricism and the Philosophy of Mind
3 Simulation and Theory-Theory


Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
3.1 Some Conceptual Clarifications

Since many similar and confusing concepts are involved here, we shall proceed to a tentative clarification. (The schema on the next page describes the taxonomy)

Interpretation concerns any activity that tries making explicit the meaning of an intentional production: an action, a sentence, a speech, a text, etc. For instance, philology is the interpretation of text, semiotics the interpretations of signs, and social science is the scientific interpretation of minded creatures and their performance. It can proceed in an engaged fashion (empathic understanding) or in a detached fashion (naturalism, logical positivism, explanation), depending whether one sees social science as primarily an empathic reconstruction of reasons or a domain of investigation similar to another. Finally, folk-psychology, the non-scientific interpretation of minded creatures and their performance, is the subject of this chapter.

Many aspects of our everyday folk-psychological practices can be studied: the cognition of emotions, personality, etc., but cognitive scientist and philosophers were mostly interested by mindreading, the ability to attribute mental states to other, to predict, explain and interpret their behavior, and to base these predictions, explanations and interpretations on attributed mental states. Again, one can adopt an engaged conception of mindreading or a detached. Proponent of the engaged conception suggest that everyday interpretation requires imaginative, rather than theoretical capacities. For Adam Smith, it is a matter of “sympathy”, which is a mix of imagination and moral concerns. For simulationists, it is a matter of empathy, i.e., recreating someone else’s thinking in order to understand her. For Jane Heal, it is an a priori or conceptual truth, while for Goldman or Gordon it is an a posteriori, empirical hypothesis.

One can distinguish three versions of the detached conception of mind reading:

1. The rationality theory: Davidson, Dennett: We use normative intuitive theories of rationality to interpret (see previous chapter). Our interpretative process as common-sense theories governed by rationality ideals.

2. The modularity theory: the theory of mind is produced by a cognitive module (Scholl & Leslie, 1999). This module is a cognitive adaptation to social life (evolutionary psychologists). [NOTE: This approach will be discussed in chapter 8]

3. The theory-theory: we have folk theories (with folk theoretical entities, folk psychological laws, folk explanation, folk prediction, etc.)
   a. External: We entertain “platitudes” about the mind such as “people prefer having what they want than what they don’t want (Lewis 1972). This common-sense theory could be systematized or analyzed. For Churchland, it is a problematic theory; it should be eliminated (1981, 2006).
   b. Internal: The theory of mind is an intuitive theory, revised according to evidence. Developmental stages are similar to paradigm change (Gopnik).

The following figure illustrates the relations between these theories:
3.2 Theoretical minds

David Lewis first articulated the view that the terms beliefs, desires, and so on function together as an explanatory and predictive theory of mind, a theory that states “the causal relations of mental states, sensory stimuli, and motor responses” (Lewis 1972: 256). Something like “When people believe that P, they take P to be true, and they are often surprised when they learned that P is false”.

Beliefs, desires, and so on are thus, according to the detached view, theoretical posits in a folk theory of mind. All platitudes use theoretical terms and determine how we use these theoretical terms in relationship with what Lewis call O-terms (others, original, older, already understood or observational): they are all implicit definition. We define intentional terms by using them with o-terms (stimuli, behaviors, etc). The implicit theory ascribes causal roles to intentional terms.

“Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses…. Add also the platitudes to the effect that one mental state falls under another – ‘toothache is a kind of pain’ and the like. Perhaps there are platitudes of other forms as well. Include only platitudes that are common knowledge among us – everyone knows them, everyone knows that everyone else knows them, and so on. For the meanings of our words are common knowledge, and I am going to claim that names of mental states derive their meaning from these platitudes.” (Lewis, 1972, p. 256)

For Lewis, folk-psychology is really folk science: the theory that gives ordinary state mental (theoretic) terms their meaning, and that specifies how to use them. Folk-psychology is common-sense psychology, it is: “folk science rather than professional science, but a theory nonetheless” Lewis, 1972, p. 250). Hence this theory must have law-like generalization such as (Churchland, 1981, p.71):

\[(x \text{ fears that } p) \rightarrow (x \text{ desires that } \neg p)\]

or

\[[(x \text{ hopes that } p) & (x \text{ discovers that } p)] \rightarrow (x \text{ is pleased that } p)\]

These sentences function like empirical generalizations that explain and predict behavior. Churchland argues that this theory is false, wrong, and should be eliminated: it is a “radically inadequate account of our internal activities, too confused and too defective to win survival through inter-theoretic reduction” (1971, p.72) On this account, folk-psychology is not an internal, cognitive process.

3.3 From External to Internal Theory

The nature of belief-desires attribution was discussed in the larger framework of theories of meaning, language use and the connection between observation and theory. Mindreading was mainly construed as dependent upon our linguistic abilities. However, with the advent of
cognitive science, a larger set of systems could be seen as having folk-psychological abilities, or a tacit theory of mind, similar to intuitive physics or intuitive biology. These systems include non-human primates and children.

Since Premack & Woodruff (1978) “Does the Chimpanzee Have a Theory of Mind?”, primatologists wondered whether chimpanzees have theory of mind (“the ability to predict and explain behavior by attributing mental states”), something they assumed humans to have. Thus they were interested in the concepts that are mastered in interpretation, not the causal mechanisms. They asked Sarah, a chimpanzee (who was trained to manipulate plastic chips and to produce strings of symbols) to choose which photograph provides the solution to a problem faced by a human experimenter (e.g. reaching for bananas). Her performance (above-chance) suggested that she had a basic theory of mind that posits intentions and beliefs. These experiments lead to many others that either support or contradict the idea that primates have a theory of mind (more details in Chapter 9).

Developmental psychology also stimulated the debate. After 3 or 4 years of age, children begin to be able to attribute false belief. They start being able to understand that other agents may act upon subjective misrepresentation of objective states of affairs; before that, they will attribute only true belief (that the attributor holds true). The standard test (the False belief test) consist in showing the children a doll (Sally, for instance) that observes an object being hidden somewhere (location 1). The doll leaves the room, while another character comes and places the object somewhere else (location 2). When the doll comes back in the room, the experimenter asks to the child “Where will Sally look?”. If the child answers location 2, she does not attribute false beliefs and thus cannot hold two perspectives in mind at the same time (hers and Sally's). She assumes that Sally knows what she (the child) knows. If the child answer location 1, then she understand that people act upon their perspective on the world (beliefs), that the character look where she thinks it is, not where it is. They predict Sally's behaviour on the basis of Sally's belief, and that is the hallmark of full-blown social interpretation and possession of a theory of mind.

Behavioral economics, the experimental study of economic behavior, also stimulate many debates in this field. First, Kahneman & Tversky (1983) cast doubt about the rationality assumption since it seems that many subjects fail to comply with basic rules of rationality, such as probability theory. For instance, subjects where presented with the following problem:

*Linda is 31 years old single outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.*

*Please rank the following statements by their probability*

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

85% of those asked chose option 2, although 2 is less probable than 1.

Tversky and Kahneman also showed that simulation is a heuristic, a tool for predicting mental states. They asked their subjects who would be more upset: a person who came to the airport 30 minutes after her flight left on time, or the person who misses her flight by 5 minutes while this same flight was supposed to leave 30 minutes earlier, but was delayed for 25 minutes? Most subjects (96%) guesses that the second person was more upset. According to Tversky and Kahneman, they simulate the passenger mental states.
Autistic children are an interesting case for the study of social cognition, since they seem to be impaired only in this domain: they do not engage in socialization, pretend play, cooperative play, imitation, communication, shared attention. Baron-Cohen, Leslie & Frith (1985) showed that autistic children are unable to pass the False belief tasks at a normal age, even when children with Down’s Syndrome can (children with intellectual impairment due to trisomy-21). Thus it was suggested that theory of mind (“social cognition” now) has a certain domain-specificity; that one can be impaired in this domain without being impaired in others. Explaining what is going wrong in the autistic child’s conceptualization of the social world suggests also how mindreading functions normally.

Consequently, the focus of attention shifted from the structure of the folk-psychological theory to its cognitive underpinning; the “theoretical” view became a thesis about the mental processing of information.

3.4 The Theory-Theory:

According to the theory-theory view (TT), folk-psychology is “an implicit theory analogous to scientific theories, and changes in that understanding may be understood as theory changes” (Gopnik & Wellman 1992). It is an internally represented theory of mind. Interpretation is thus a theoretical inference that use knowledge about psychological generalizations: we know that Pete, when he bought apples at the fruit store, wanted to have apples and knew that he could have some at the fruit store, because we have, in our intuitive theory of mind, a rule such as the “central action principle” (CAP): If somebody desires x and believes that A-ing is a means of achieving X, then, ceteris paribus, he will do A” (Stueber 2006, 110)

A theory is explanatory (why A did X?), predictive (what will A do?), systematic (A does X because people generally....), coherent, abstract (beliefs, desires, etc. explains A's action), defeasible (can be falsified/verified, revised, abandoned, etc.). It posits theoretical entities (e.g. gravity) and laws (e.g. Newton's). If we have, or acquire, a folk theory of mind, we have representations of these entities/laws. Theories also produce interpretations (A thinks that the juice is in the refrigerator) of evidence, not just descriptions and generalization.

According to TT proponents, children acquire the theory of mind progressively and through a process that mimics scientific change: they acquire a more complex theory that supersedes the precedent. The acquisition passes through 3 stages (Gopnik & Wellman 1992)

At 2-year-old, the child has a basic, incomplete theory that does not posit mental representational states. She has a basic desire-perception theory: desire (drives) and perception (awareness of objects) are the basic categories. At 3-year-old, she possesses an intermediate understanding of the existence of representational states as auxiliary hypotheses. She starts understanding false representational states (pretenses, dreams, and images), start using mental vocabulary, but see mental states as having a direct causal links between objects and agents. By 4-years-old, she reorganizes her central explanatory theory that becomes a genuine belief-desire psychology and develop a “representational model of mind” mentality is mediated by representations.

She thus lives a small-scale “Copernican revolution”, from a non-representational to a representational theory. First the theory is protected from counter-evidence, but as they accumulate they cause a theory change. It adjusts pre-existing concepts (drives-perception) and
reformulates them in a wider, systematical structure, the belief-desire psychology. It is a genuine theory because “it involves general constructs about the mind that go beyond the focal evidential phenomena” (Gopnik & Wellman 1992, p.153): children make predictions, can see these predictions as wrong, explain behavior and interpret behavior with their theory. Cognitive development is thus strongly analogous with scientific evolution:

“we would say, not that children are little scientists but that scientists are big, and relatively slow, children” (Gopnik & Wellman 1992, 168).

“Everyday cognition, on this view, is simply the theory that most of us most of the time have arrived at when we get too old and stupid to do more theorizing (...) We might think of our enterprise as scientists as the further revision of the theory by the fortunate, or possibly just childish, few who are given leisure to collect evidence and think about it. (Gopnik & Meltzoff 1997, p. 214)

3.5 Simulating minds

The first wave (50's-80's) of cognitive science was mostly interested in problem-solving, expert systems and higher cognition. Since the 80's, many researchers (from robotics, linguistics, neuroscience, AI, psychology, philosophy, etc.) have suggested that the classical picture of the mind as a rational deliberator was wrong. According to the new approaches, cognitive processes are not limited to the symbolic processing of internal information structures, but implemented in various sensorimotor processes (motricity, perception, emotions, coordination, imagery, emulation, simulation) and various substrates (members, bodies, artifacts, environmental regularities). The major theme of these new approaches is the embeddedness of cognition: in the body and in the world. If one can, conceptually, distinguish the brain from the body and from the environment, a dense and continuous flow of information binds the three together. Whether it is problem-solving or perception, decision-making or memory, many cognitive processes “piggyback” on lower-level processes. High-level processes, such as voluntary action, guidance or categorization, can be seen as complex redeploys of sensorimotor processes. The constant sensorimotor dynamics, the role of action in perception and learning, the agent/environment coupling and collective intelligence are cognitive determinants which one must take into account to adequately describe natural cognitive systems and to build artificial cognitive systems. If this picture of the mind is correct, if we are “better at Frisbee than at logic” (Clark, p.60), then our scientific and philosophical picture of the mind should acknowledge that.

Simulation theory represents a departure from the founding picture of an “intuitive scientist”: it gives a new importance to skill, imagination, perception, and emotions. Non-linguistic knowledge has a place in cognitive science and philosophy of mind. Philosophically speaking, the simulation theory proposes a conception closer to empirical facts than other approaches (at least in philosophy) that were more interested in norms of interpretation and not the actual practice of interpretation.

Simulation theory denies the centrality of theorizing in interpretation, prediction and explanation. It is about “Understanding other minds from the inside” (Heal, 1998). It emphasizes
the first-person perspective in understanding the mind, and the use of models, instead of theory. Certain theorists put emphasis on the use of models (Goldman), others on the first-person perspective (Stueber). It can be construed as an a priori thesis or an empirical hypothesis.

3.6 A Posteriori Simulation

Goldman (1989) criticizes Theory-Theory (TT) because of its vagueness and inaccuracy: psychological laws are vague (full of ceteris paribus clauses) and often inaccurate (no distinctions between occurrent beliefs/desires and dispositional beliefs/desires. When philosophers articulate “mentalistic platitudes”, they are not really making a theory explicit, but creating these laws. Another problem is that in development, children engage early in interpretive practice but were never exposed to anything like a theory. If they acquire it by their own, then the problem is that “all children miraculously construct the same nomological principles” (1989, 168).

Attributing beliefs and desires is rather an imaginative process: in interpretation, we consider what “we should do if we had the relevant beliefs and desires.” (1989, 168). In playing chess (cf. Gordon, 1986), we imagine what the other player would do. The only theoretical knowledge is the knowledge of chess rules, not propositional knowledge about human cognition; it is more a skill, a “knowing-how”. Through simulation we can predict, but also explain and “retrodict” i.e., reconstruct an action after the fact. We use it to infer mental states from mental states (in chess: inferring from the actual mental states of the other player to her future mental states), but also for inferring actions from mental states: what one will do given her actual mental state. Of course, the simulation can be adjusted to the situation, or the target agent (a chess master, for instance). It is not a perfect, but a central interpretative strategy. Inductions or psychological regularities can supplement simulation.

The central tenet of ST is thus that contrary to other domains, interpretation requires an inner imitation, the use of our own emotional and motivational and cognitive resources. Our relations with other persons is different from our cognitive grasps of other domain because all humans are similarly-minded creatures: a rock or a fly is not, thus it requires other mental resources. We use our imaginative capacities to see the world from another's perspective without seeing as true: we entertain “make-believes”. We already use our imagination in many occasions to simulate our possible or future self (what would happen if I lose my job?); in order to simulate another creature, we must share a certain set of mental capacities (it is hard to imagine what it's like to be a bat (Nagel, 1974)). Since we know that other humans have similar mental capacities, I am able to use myself as a model of someone else provided that I see her as a person. What warrants my simulation of myself is the belief that me-now and me-next week share certain mental dispositions, a similar responsiveness to reasons, etc. This also warrants my simulation of someone else: I know how where Sally will look because I can pretend to be Sally, to see things from her perspective and entertain the belief that the object is in location 1 while at the same time quarantining this belief. I share certain cognitive similarity with Sally and this is why my simulation can work and why it cannot work (at least not completely) with bats. It is informative in the first case, but not the second.

An advantage of ST over the rationality approach is that I can simulate inconsistent beliefs; I can simulate the “intuitive pull” of the conjunction fallacy, even though it is irrational. It also has an advantage over the Theory-theory approach. For instance, in explaining humor: we
know how to make jokes without having a theory of humor. Note that theoretical knowledge is completely absent, but in many situations, we know how to make someone laugh because we can simulate their reaction to our jokes.

Simulation process, according to Goldman goes through 3 steps:
1. Matching: adopting another's perspective
2. Simulation: thinking from another’s perspective
3. Attribution: back to self-perspective, using the knowledge from the simulation for interpreting the other

(There can be theoretical knowledge involved in the matching and attribution process, but the simulation does not require at theory of mind.)

Thus simulation works if there is 1) a basic similarity in mental structure and 2) if I am able to adjust my judgment to certain relevant difference between my mental states (beliefs-desires) and the other person's states (Stueber, 2006p. 112)

"The simulation hypothesis assumes that the interpreter tends to impute to the interpretee the same fundamental categories as her own, or at least the same basic category-forming (and proposition-forming) operations. She also tends to project the same basic belief-forming processes”. (Goldman, 1989, p.180)

Thus in simulating, I am not consulting a theory about human cognition and human behavior. I can use some theory that I attribute to the target agent (if I try to interpret a lawyer), but the theory is not a theory of mind. I use my own cognitive process for reconstructing the reasons of an action or for predicting what would the target agent do. (All this process, according to Stueber, is governed by the rationality assumption).

The process does not need to be phenomenologically salient: it is a skill, something we do naturally, without thinking (such as walking, breathing, etc.). Moreover, simulation can develop into scripts, schemas (e.g. how the Prime Minster talks in Parliament, How to order a meal in a restaurant). Cognitively speaking, it is also a thesis about an overlap between mechanisms for understanding others and the mechanisms involved in our own mental processes. Simulations execute psychological models (based on similarity) instead of executing deductive translations between propositions.

Many proponents argue that we simulate from the personal-level perspective (e.g. Gordon, Stueber). Goldman sees it as an inference “from me to you” instead of a re-enactment. Moreover, Goldman rejects the ideal of rationality in the simulation.

3.7 A Priori Simulation

For Jane Heal (1998, p.477) simulation is an “a priori claim about the relations of certain personal level cognitive abilities, [not an] a posteriori hypothesis about the workings of sub-personal cognitive machinery”. She calls it “co-cognition”. It is, as Stueber puts it, a first-person perspective activity:
"in simulating another person I situate myself toward the world in such a way that I start thinking about it very similarly to the person whose mind I am trying to read” (Stueber, 2006, p.123)

For Heal, it is an a priori truth (as in “deeply embedded in our world view”) that to think that A believes that P, we also have to think that P: we” co-cognize”. The very nature of thought-as a truth-oriented perspective--entails that attributing to someone the thought that P, involves that I think about P as the other person think about P, because my thinking about P and thinking that someone else thinks about P are connected subject matters. For example, photographs of vegetables and vegetables are not independent subject matters; you cannot grasp the first one if you do not grasp the second one. Similarly, you cannot grasp “A believes that P” is you cannot grasp what believing that P is. Thus grasping someone else's perspective depends upon having a perspective, and using this perspective to arrive at the same conclusion that the target agent does.

Moreover, I have to be able to understand (comprehend) P before attributing that P. If I don't speak French, I cannot understand “le chat est sur le sofa”, but I can understand that Marcel from Paris and Maurice from Toulouse have the same thought when they think “le chat est sur le canapé”. I can attribute this mental states, but I would not understand it. I cannot simulate anything (for instance, what they could infer from that). In order to interpret someone as thinking that “the cat is on the sofa”, I must be able to think that “the cat is on the sofa” as if I think it myself. Interpreting another person’s thoughts or utterances requires personal understanding.

In sum, if you are a thoughts-entertaining creature, it is an a priori truth that interpreting another thoughts-entertaining creature requires a personal-level and egocentric simulation of the thoughts of the target creature. If someone believes that P1, P2, P3....Pn, and I wonders whether Q, I cannot guess what she will conclude if I cannot think about the states of affairs described by P1, P2, P3...Pn. I thus co-cognize (re-construct the rational pattern); I am not applying a theory of Ps, or a theory about inferential practices.

3.8 Comparisons

TT proponents recognize that imagination is important in interpretation, but argue that interpretation require intuitive theoretical knowledge. Moreover, they suggest that ST cannot account for cognitive development. At a certain moment, children have certain concepts and not others, make certain errors but not others (e.g. in reporting past mental states, 3 years-old are good at reporting past desires, but not past beliefs). According to Gopnik and Wellman, ST cannot explain that because the difficulty between easy and hard attributions does not map onto self vs. other attributions. Three years-olds fail to report past belief because they lack the concept of belief; it is a theoretical problem, not a simulatory problem.

Why do children use beliefs and desires (and not pain, fears, imagination, etc.) in action interpretation? TT, claim its proponents, can explain this centrality: beliefs and desires are the basic concepts in the child's theory of mind. The 3 years-old do not make a difference between knowing and guessing, between sources and probability of knowledge. ST has no rationale for this fact, while TT can explain it as a lack of a complex causal-explanatory structure.

TT, ST, RT can be compared with regard to their conception of the Central action principle (“If somebody desires x and believes that A-ing is a means of achieving X, then, ceteris
paribus, he will do A”): for TT, it is a theory, for ST it is an implicit know-how, and for RT it is an a priori condition.

Finally, a last distinction between ST and TT: the first one is information-poor while the other is information-rich. This is evident when we compare their claim on several matters:

Prediction: how do you predict you opponent in a chess game? a theory of mind plus a theory of chess-playing, or do you imagine what you would do in his position?

Pretend play: does it require a theory of mommy or the doctor, or a capacity to simulate another's perspective?

Development: Is development a progress in imaginative capacity or a conceptual change?

Mechanism: Do we use the same mechanism in action/speech production and understanding, two different mechanisms? (the theory of mind is not the mechanism that produce action and speech)

Finally, ST and TT can be compared along those lines:

<table>
<thead>
<tr>
<th>Cognitive mechanism</th>
<th>Simulation:</th>
<th>Theory-Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simulation skill – inner imitation (knowing-how)</td>
<td>Intuitive theory – folk science (knowing-that)</td>
</tr>
<tr>
<td></td>
<td>Imagination</td>
<td>Reasoning</td>
</tr>
<tr>
<td></td>
<td>Replication, re-creation</td>
<td>Observation</td>
</tr>
<tr>
<td>Perspective</td>
<td>Egocentric, Subjective: “replicating” the target’s thoughts</td>
<td>Allocentric, Objective: observing the target’s thoughts</td>
</tr>
<tr>
<td>Prediction</td>
<td>Forward simulation: Imagining the target perspective, mental states, and motivations and the actions that would follow</td>
<td>Deriving a prediction from folk psychological law-like generalizations cum knowledge of the situation and the agent.</td>
</tr>
<tr>
<td>Explanation</td>
<td>From an action A, simulating postulated past mental states, and see if a) simulation leads to A, and if b) simulation leads to choosing A</td>
<td>Subsuming the agent’s action under folk psychological law-like generalizations cum knowledge of the situation and the agent.</td>
</tr>
<tr>
<td>Interpretation</td>
<td>Feigning beliefs and desires. Ability to simulate, to reconstruct a thought sequence</td>
<td>Deducing beliefs and desires. Ability to subsume behavior under the intuitive theory</td>
</tr>
<tr>
<td>Mental representations</td>
<td>Models (similarity-based) Practical knowledge</td>
<td>Theories (symbolic representation) Abstract knowledge</td>
</tr>
<tr>
<td>Role of the “Central Action Principle” (“If somebody desires x and believes that A-ing is a means of achieving X, then, ceteris paribus, he will do A”)</td>
<td>Implicit in the simulation process</td>
<td>Basic postulate of the theory</td>
</tr>
<tr>
<td>Information required</td>
<td>Information-poor</td>
<td>Information-rich</td>
</tr>
<tr>
<td>Development</td>
<td>Progress in imaginative capacity</td>
<td>Conceptual change</td>
</tr>
<tr>
<td>Mechanisms of action production/understanding</td>
<td>Important overlap: redeployment of cognitive activity in “pretend” mode</td>
<td>Separate: the intuitive theory does not produce behaviors</td>
</tr>
<tr>
<td>Domain-specificity</td>
<td>Simulating other minds is a process dedicated to social understanding</td>
<td>Explaining other minds is similar to explanation in other domains (e.g. folk-physics)</td>
</tr>
</tbody>
</table>
3.9 References


4 Embodied Phenomenology and Narratives


Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
4.1 The Critique of Folk-Psychological Reason

The contemporary discussions about the nature of folk-psychology (that culminated in the simulation theory vs. theory-theory debated) are the offspring of several lines of research. The first one is the Wittgenstein-Anscombe-Davidson questioning about the nature of action descriptions. After Davidson, it was common to see belief and desires as propositional attitudes that rationalize and cause actions. The second one started with Sellars, Lewis and Quine, and then with Dennett, Fodor and Churchland: inquiries about the function of intentional states ascriptions. Now beliefs and desires are commonly thought of as mental representations—either as real ones (Fodor), useful fictions (Dennett) or useless fictions (Churchland)—used in prediction and explanation. The third one is constituted by all cognitive scientists interested in developmental, comparative and abnormal psychology: the presence (or absence) of a 'theory of mind' in primates, children, autistic children, etc.

Consequently, the standard account of folk-psychology became something like:

*folk-psychology is the commonsense, intuitive framework we use to predict and explain actions; it consists in attributing propositional attitudes (mainly belief-desires).*

Any philosophy textbook presents a similar definition. Against the once dominant theoretical view, the simulationists insisted that in interpreting, we are not applying theoretical knowledge, but using our imagination and our own cognitive processes to make belief-desires predictions and explanations. The basic assumptions were not challenged.

As this chapter and the following ones will illustrate, many disagreements with the standard propositional-attitude belief-desire psychology are possible. Many of its theses can be disputed:

The content-ascription thesis: do we always ascribe content in interpretation?

The content of folk-psychology (when we attribute content, is it propositional attitudes? and always beliefs and desires? what about, e.g. character traits?)

The function folk-psychology (are we only predicting and explaining? what about evaluating, judging, justifying. And when we predict and explain is it always based on belief-desires attribution?)

The psychological role of folk-psychology (what if belief-desire psychology is separate from actual interpretation?)

The unity of folk-psychology (are all belief-attributions identical, or are there many kinds of beliefs attribution?)

The evidential basis for philosophical accounts of folk-psychology (why using only conceptual analysis and a priori thinking?)

The neglected role of phenomenology, moral judgment, non-conceptual content, narratives

The interpretive perspective: why only third-person or first-person? What about second-person?

One way to challenge the standard account is by arguing that interpretation is not a matter of simulating or theorizing, but an activity that, most of the time, recruits basic sensorimotor capacity and not explicit thought. This is Shaun Gallagher’s embodied phenomenology position.
Daniel Hutto also challenges the standard account: folk-psychology is not primarily an interpretive mechanism that manipulates beliefs and desires in order to explain or predict behavior: it is primarily a narrative practice. Folk-psychological discourse is essentially about the construction of reason-based narrative that figures in dialogues and conversation.

4.2 Phenomeology and the limits of “theories of mind”

According to TT, we use a theoretical stance to interpret, while, for ST, we use our “own experience as the measure of everyone else's”. But none of them, according to Gallagher, is the primary way in which we interpret others (Gallagher, 2001). Both share the same “mentalistic supposition”: to know something about other minds is to know their beliefs, desires etc. We either simulate their belief-desires (ST), or derive them from folk-psychological laws (TT)

In both case, we allegedly use conceptual knowledge: we need the concept of belief-desires to simulate/theorize about belief-desires. Even if these inference are implicit, they are informed by concepts: “To discover a belief as an intentional state even in myself requires that I take up a second-order reflective stance and recognize that my cognitive action can be classified as a belief (...) [it] requires something like a reflective detachment from my phenomenal experience “(Gallagher, 2001). ST and TT require always a certain detachment, a representational manipulation of theoretical knowledge or a representational simulation

ST and TT may capture what is going on when we talk with someone else about a third person: we may attribute or simulate the third person mental states. For Gallagher, however, these processes do not capture the nature of the interaction with our interlocutor.

“in a second-person conversational situation, although we may indeed tacitly follow certain rules of conversation, our process of interpretation does not seem to involve a detached or abstract, third-person quest for causal explanation. Nor does it seem to be a theory-driven interpretation that takes the other person's words as evidence for a mental state standing behind what he has just said. “ (Gallagher, 2001)

In fact, Gallagher challenges the very pre-suppositions of ST and TT: that, in interpersonal understanding, we posit a “theoretical entity called a belief and attribute it to [the interlocutor]”. He challenges what we could call the bi-Cartesian account of communication: Instead of construing communication as a process where there is first an exchange of information, that is processed by theoretical or simulation mechanisms and afterwards expressed in speech or behavior, Gallagher argues that the understanding happens in the process of communication. We do not understand someone first as saying something and then understanding the “belief in her head”. We rather understand--normally--when the person say something. We understand each others in a nonmentalizing way (neither reasoning (TT) nor imagination (ST): we use instead “embodied practices—practices that are emotional, sensory-motor, perceptual and nonconceptual” (Gallagher, 2001, p. 81).

We use third-person or conceptual knowledge only when regular, situated action does not follow its regular course. Normally, our understanding of the mental is just like our understanding of everyday objects we may use or everyday situations we may be in without any theory, knowledge or explanatory entities. To use Heidegger's example, the carpenter does not need conceptual knowledge of the hammer when she uses it: only when the hammer breaks, or when something goes wrong with it does the hammer become an object of thought. Most of the
time, the hammer become an extension of the body and is “absorbed” in the action. Similarly, social understanding appears phenomenologically as immersed in pragmatic actions such as communication. We rely on theories and simulation only when social interaction does not follow its normal course.

Gallagher gives this example. Imagine you hear this conversation:

**Woman:** I'm leaving you.

**Man:** Who is he?

We could explain our understanding of the man's reaction by invoking his beliefs and desires, social scripts, schema and scenario that explains his conclusion. We could, but the real question is: when we hear this conversation, are we naturally prompted to explain it, or aren't we rather directly understanding his state of mind, or more precisely “comprehending it in an evaluative way” (Ibid.)? His thought is already given to us in his speech, so that we do not need to be engaged in theorizing/simulating. Social understanding is centrally dependent upon non-theoretical capacities: such as the recognition of movements, gestures, postures, facial expressions, etc. It is an “embodied” practice, developmentally and psychologically primary. Simulated and theoretical understanding are complex refinements of this capacity.

We do not need simulation or theory when we chat about our day at work: the interlocutor's states of mind are not really objects of thought, of simulation or of theorizing, but part of the communication process. What we do in the course of interaction, is not ascribing beliefs or desires so that we can re-construct the interlocutor mental states but, according to Gallagher, evaluative: we judge, agree or disagree, approve or disapprove, etc. Interpreting someone does not happen from a detached perspective, but in a situated, engaged perspective. It requires pre-theoretical abilities already present in children (before the child use mentalistic vocabulary). Similarly, we do not need to posit abstract or functional states beliefs and desires or propositional attributions “behind” overt behavior and speech, since all we need to know is in the behavioral or linguistic expression: gestures, emotions, tone of voice, stare, etc. Simulationists tend to define simulation as a modeling process over which we have a certain control: we feed our decision-making process with pretend beliefs and desires, then use the output of this process to predict or explain someone's behavior. But at the personal level we, as subjects and agents, do not do anything: we just see that someone is happy to see us when she smiles and says “It's good to see you!” We do not have any control over our automatic perception of her intention and over what she meant:

*The other person has an effect on us (...). This is not a simulation, but a perceptual elicitation. It is not us (or our brain) doing it, but the other who does this to us (Gallagher, 2007, p. 73)*

Thus interpretation is not an instrumental action like modeling.

While ST and TT suppose that we predict and explain each other, phenomenological accounts see social interaction as less theoretical, more situated: in most of our interactions we judge what the other means, or how we should behave instead of simulating/theorizing about beliefs and desires. Social interaction occurs with a high degree of engagement: it is more a pragmatic action than a conceptual reflection. But then, when we take a more reflective stance, aren't we using theory and simulation? According to Gallagher and others (e.g. Hutto), no.
Understanding is not primarily about theory or simulation, about speculative causes. We instead use narratives to understand reasons.

4.3 Folk-psychological narratives

Daniel Hutto's *Narrative Practice Hypothesis* (NPH) challenges the supposed primacy and pervasiveness of folk-psychology. Social cognition, most of the time, does not “involve the manipulation of representations by inferential operations (let alone representations of propositional attitudes).” (Hutto, 2007, p. 116). We use scripts, heuristics and fast embodied recognitional capacities to navigate our social environment. We rarely need to “read minds”. Moreover, FP is not fundamental in social interaction and interpretation. Reason-based understanding is in fact limited to normal adult human beings. Yet even this claim should be restricted: FP schemas are not cultural universals. Although people make sense of each other, they do not always use the belief-desires framework: certain cultures favor trait-based or supernatural explanations (e.g., Lillard 1997). “Folk-psychology” might be in fact the European-American version of social interpretation. Hence FP is a restricted competence.

Hutto does not challenge that we use the belief-desires framework, but challenges the idea that we use it for speculating about other people mental states. We rather use it when we produce narratives that make sense of their action.

“asking the other for their reasons is vastly more reliable than trying to determine why they in fact acted as they did from the distance of a third-party spectator (...). It is only in second-personal contexts that we confidently obtain true folk psychological explanations, by and large, as opposed to speculating about merely possible ones” (Hutto, 2007 p. 118)

Instead of theorizing or simulating beliefs and desires, folk-psychology is a narrative practice. When we reflect upon others or us in a more detached way, we “generate a narrative framework that would facilitate our understanding of them” (Gallagher, 2007, pp.63-64).

According to Hutto, FP is learned in, and is used to produce, a certain type of stories: folk-psychological narrative, where agents act for reasons.

“it is through direct encounters with stories about reasons for acting, those supplied by responsive caregivers in interactive contexts, that children become familiar with (1) the core structure of folk psychology and (2) the norm-governed possibilities for wielding it in practice (i.e. learning both how and when to apply it).” (Hutto, 2007, p. 117)

Thus the content and the use of the belief-desires framework (reason-based explanations, interpretation, prediction, etc.) is developed through stories. It has a socio-cultural basis, not an innate one. Narrative practices are not folk-psychological modules (cf. chapter 8) or processes that we acquired in our evolutionary history: they are rather recent, human linguistic activities. Our ancestors had “interactive, imaginative and mimetic capacities”(p. 119), but not folk-psychological ones.

We encounter folk-psychological narratives (stories about agents who act upon reasons) mainly in dialogue, where people produce narrations. Thus the primary function of FP is to produce and understand narratives. When we explain, predict or interpret people, we apply our narrative competence. We tell stories about others and ourselves to others. We make sense of others not in applying a theory or in simulating their mental states, but in interactive linguistic
exchange, in asking them about reasons, and listening to their narrative. We are not spectator, but actor in folk-psychological practices.

“it is these second-person deliveries – these narrations – that do the heavy lifting in enabling us to understand and make sense of others with confidence.” (Hutto, 2007 p. 118)

Hence our folk-psychological competence is principally located in second-person dialogue and conversation, not in third-person or first-person observations. However, theoretical and simulation competence can supplement our narrative activity in situations where we are spectators. We sometimes have to speculate about mental states—especially when we don’t have enough information about the target agent—and thus have to use our theoretical and simulation heuristics; these “Holmesian” heuristics, as Hutto call them, are not essentially what FP is about: our folk-psychological competence is centrally narrative and primarily applied in interactive context.

Thus Hutto present both an account of the etiology and function of FP: it is acquired and applied primarily in second-personal engagements. FP is thus intimately tied to the possession of a language; like Davidson, Hutto argues that interpretation requires linguistic abilities. In order to attribute reasons, one must have the ability to represent representations (otherwise the agent is a Rylean ancestor). Having propositional attitudes implies a three-place relation between a thinker, a sentence, and the state of affairs that the sentence refers to. Sentences have the right kind of logical, semantic, compositional, computational and inferential features that reasons require. Having a propositional attitude requires having an attitude toward a certain sentence. To manipulate sentence, one must possess a language: to think with sentences, one must think in a language. With language comes “supermental thinking”: it is a reconfiguration of mental habits, disposition and capacities.

Attributing a propositional attitude thus involves attributing to someone else an attitude toward a sentence. As in Sellars's Rylean ancestors myth, reasons and their logical articulation are modeled after speech acts. Linguistic beings start using the belief-desires framework when they attribute 'inner episodes' of practical deliberation and 'inner sentences' similar to public episodes of practical deliberation and public speech. A compositional language is thus necessary for practical reasoning; the capacity to describe practical reasoning also requires language. Thus folk-psychological practices are likely to be not older than 35–40,000 years. They began when linguistic beings endowed with practical reasoning abilities (e.g. Sellars's Jones) began to be able to narrate their actions:

“The authors of certain actions would have given accounts of the plans they constructed based on propositional beliefs and desires, at least” (Hutto, 2007, p. 121)

Understanding reasons thus require inter-subjectivity as it is displayed in dialogues. Folk-psychology begins when the second-personal perspective on reasons, when they are narrated in stories, are employed in third-personal speculations. It is thus trough stories that FP framework develop, not from inherited mechanisms.

FP is a multi-layered ability acquired progressively in development. It starts with embodied skills that do not require mental content ascription (e.g. joint attention, eye-tracking, etc.), but are necessary for reasons ascription. Then children have meta-representational abilities: they can attribute beliefs-as-subjective-perspectives, but not beliefs-as-reasons-for-action: the
child uses FP abilities when she understand that beliefs are logically articulated with action, that A does X because he believes that P: P is not just a perspective from A's point of view (as in the false belief task), it is also a reason for doing X. Children must understand the elements (beliefs, desires, etc.) but also how they are appropriately combined. Folk-psychological narratives about reasons provide “instruction” for mixing meta-representational “ingredients”. Other embodied and imaginative skills are recruited for understanding narratives (as stories and acts of narration, stories and story-telling). Through autobiographical accounts, gossip, oral transmission, fairy tales (Little Red Riding Hood), myths, etc., stories familiarize children with the content and norms of the folk-psychological framework

NPH accounts for the fact that mentalistic concepts are always located in a structured network (as Lewis argued), but does not imply that FP is a theory. Contrary to external theoretical accounts, learning FP, pace Lewis, is not learning a theory, but learning conventional practices (how to use words, concepts, stories, etc.). It is similar to a theory only because it has the same inferential structure and holistic semantics. But it is not a theory: a theory is a third-personal, observational practice. Folk-Psychology is a narrative practice.

Contrary to TT, it does not explain the development of FP as a development of theoretical capacities, but practical capacities: embodied skills, scaffolded tools, and sociocultural practices. Simulation and theories are employed only when we speculate about possible reasons for actions. In conversation and folk-psychological narrative, we don't speculate, theorize or engage in pretend thinking: we just talk or listen. But these are heuristics, peripheral uses of FP capacities.

4.4 References


5 Experimental Approaches to Folk-Psychology: Moral Judgments and Pluralistic Accounts

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
5.1 Folk-Psychology, From Intuitions to Experimentation

According to what we called the *standard account* in the last chapter, Folk-Psychology is a set of mechanisms or a body of knowledge that is involved principally in two cognitive operations: prediction and explanation. *Prediction* is an inference from mental states (beliefs-desires) to action, while *explanation* is an inference from action to mental states. This supposition—common in all major philosophy textbooks—has always been undisputed. Recently, however, experimental psychologist and philosophers undertake a more systematic approach of FP.

Malle and other social psychologists that try to decipher the folk concept of intentionality suppose three ‘layers’: a conceptual framework, a set of psychological processes and linguistic forms. The suggestion is that we all have a conceptual framework, akin to a “deep grammar” for social explanations. This framework is then expressed in linguistic forms by a set of psychological processes that govern the construction of explanations. Malles and his collaborators found that the conceptual framework, its psychological processing and its linguistic expression is relatively similar from one individual to another (Malle, 2001, 2007). First, almost everybody who agrees whether an action is intentional or not, rely preferentially on causes to explain the former and reasons to explain the latter. About 70% of the intentional actions are explains by primary reasons: beliefs, desires but also valuings (e.g. “she got home late because she liked the show”). When primary reasons are not evoked, subjects use either a causal history of reasons (CHR) explanation or an enabling-factor explanation. The first one explains why a person decided to do X not because of her beliefs/desires, but because of factors that bring about reasons to act: for instance, “she comes from a respectful culture”. The enabling-factor explanations cite—after the action is performed—the condition that made its performance possible without referring to the agent’s intentions or motivations (e.g. “she had two week to prepare the talk”). Here is how Malle et al. represent the folk concept of intentionality, based on their experiments:

![Diagram of Folk-Psychology](from Malle, 2007)

One the one hand, empirical research indicates that philosophers were right to see belief/desire as having a central role in FP (they are involved in the majority of explanations). On
the other hand, they may have overestimated the role of belief/desire attribution: folk-psychology seems to be a richer framework. Thus certain philosophers now propose a pluralistic account of FP.

5.2 Pluralism in Folk-Psychology

As Goldie (2007) argues, belief/desires explanations are rather thin: spelling out the mental content that explain an action is not really informative or explanatory. It is rather a re-description of the action. To rationalize an action (in the thin sense) consists just in showing how an action makes sense for someone, but there are many ways in which the action can make sense. One only has to look how we can make different, but all meaningful, predictions of someone’s behavior.

Common-sense psychology, according to pluralist accounts (e.g. Andrews 2007, Goldie 2007) produce thick, rather than thin, explanations. Most of them are richer than beliefs/desires. A rapid search on Google using the expression “He did that because” output different styles of explanation:

- He did that because of his friendship with the minister [valuing, causal history of reasons]
- He did that because he loves you. [valuing, motive]
- He did that because the consensus had moved [external factors]
- He did that because he believes it and he did that because he wants to say it for her too. [beliefs-desires explanation]
- He did that because of misconduct allegations [external factors]
- He did that because I’m disabled [internal factors]
- He did that because he had that quality of being genuine, of being authentic, of being unconditional and having a great zest for life [character trait]
- He did that because he wanted the marriage to work [desire]

As these examples (and Malle’s research) indicate, folk psychology is pluralistic and not limited to the attribution of propositional attitudes. We use many different concepts in explanation and prediction. We use factors such as states (depression, drunkenness, tiredness), emotions (she was angry, happy) moods (these days, he is tense, irritable, sad). Although these factors are not reasons (they are not located in the “space of reasons”), they do affect reasoning and thus can explain how someone decided to do something. We also use stereotypes (gender-, sex-, job-related), rules about what one should do given a certain situation or generalizations over past behaviors.

Motives are also often used in explanations. Motives explain an action by showing how something is desirable for the agent, how the object of desire is desirable. The motive does not need to be psychologically occurrent. One can be involved in a social practice without thinking about the motive (e.g. satisfying clients). Motive explanations are more general than B/D: they do not refer to “what's in someone’s mind”, but how a certain disposition is operative in a certain situation.

Another important category are personality traits. While motives indicate the effectiveness of disposition, traits are more general dispositions, or more precisely stable
properties that describe behavioral disposition. They are spontaneously used in descriptions, prediction, explanation, evaluation, interpretation, etc. Traits are thought to be less predictive than attitude (traits are biased: the fundamental attribution error biases people to neglect the influence of the situation on behavior) but we are generally biased in overestimating our predictive abilities and in confirming our hypothesis, so we don't notice faulty predictions (Andrews, 2007). Moreover, trait-based predictions are often local and contextualized, therefore we don't have many counterexamples (only when we generalize: we might be good at predicting a co-worker behavior at work, but not on a trip). Also, since the predictor is often part of the predicted situation, s/he can influence the predictability of the situation (e.g. the co-worker might always behave similarly with you). Traits are one heuristics among many others.

But is a trait just a complex of beliefs and desires? One might say that generous people have a desire to help others and belief that doing so-and-so will contribute to helping them. If that is the case, then trait-based predictions are just BD-predictions. But children use trait attributions only after they use BD attributions; hence trait attribution is a different capacity. Yet this does not show that this does not come from a generalization over BD attributions. However, although autistic children are unable to attribute mental states, they can learn (through Social Stories Therapy) to make predictions based on traits. The child can associate smiles with 'happy', and 'happy' with certain types of behavior. She can make prediction without attributing BD.

This 'folk theory of behavior' understanding is not restricted to autistic children; normal subjects also use it. Thus trait attributions are not reducible to belief attributions, and trait explanations are not mentalistic explanations in essence.

5.3 The Experimental Philosophy of Folk-Psychological Judgment

One of the consequences of the standard account, even if we consider the richer version provided by experimentalists, is that moral judgment allegedly takes place after folk-psychological judgment. First, we think “X does A”, then we judge that “X is blameworthy for doing A”. Experiments conducted by Joshua Knobe and other “experimental philosophers” (philosophers that use empirical method to evaluate philosophical or commonsense claims1) showed that in many cases, people use a moral judgments to generate a FP judgment.

For instance, one of the first finding is that the moral status of a behavior can have an impact on whether or not people regard it as intentional (Knobe, 2007, p. 158). Take the following story

“The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will harm us increase profits, and it will also harm the environment.’ The chairman of the board answered, ‘I don’t care at all about harming the environment. I just want to make as much profit as I can. Let’s start the new program.’ They started the new program. Sure enough, the environment was harmed’”

1 The “X-Phi” community has a lively blog: http://experimentalphilosophy.typepad.com/
Most people claim that the chairman intentionally harmed the environment. If “harm” is replaced by “help”, however, people claim that the chairman unintentionally help the environment.

Since the two cases have the same belief-desire structure, both actions should be seen as intentional, whether it is right or wrong. It turns out that in the “harm” version, most people (82%) say that the chairman intentionally harm the environment; in the “help” version, only 23% say that the chairman intentionally help the environment. This effect is called the “Knobe effect”, because it was discovered by philosopher Joshua Knobe. The experiment was also replicated with Hindi speakers—and led to similar results—and children. In the latter, the story was simplified. Children (2-,3-4-years old) were presented with a situation about a child A bringing a dog home, and knowing that it would make his friend B happy/upset. They were then asked “Does [A] make [B] happy/upset on purpose? When it upsets B, 4-years-old judge that A did it on purpose (but not younger children). This is a significant result, because this is the age where children usually acquire a full-blown theory of mind.

Other research led to similar results with the concept of valuing, that seems to be sensitive to moral consideration. Typically, philosophers and common sense distinguish desires from values. There is a difference between wanting a cigarette and valuing it: the happy smoker desires and wants it; the smoker who tries to quit smoking desires a cigarette but may not value it. It is hard, however, to sharply separate one from the other. Knobe suggested that valuing might be a prototype concept, that is, a concept that is not characterized by individually necessary and conjointly sufficient conditions (e.g., to be a grand-mother is to be the mother of the mother), but a prototype: x is C iff x has some of the features of a prototypical C (the typical car has 4 wheels, 4 places, etc.). Thus the prototype of valuing would be something like (i.e. a typical case of valuing will conform to most of these):

1) the agent has a conscious belief that o is good
2) the agent is motivated to promote o
3) the agent experiences guilt when she fails to promote o in circumstances where she could have
4) the agent has a second-order desire for o (i.e., a desire to desire o)

If there is a “Knobe effect” for valuing, we may be attributing valuations to people not only when we think they esteem something, but when we think that it is a morally good object. Maybe we consider that the heroin addict does not value heroin because it is really bad for him. If that is the case, then there is another feature of the valuing prototype:

5) whether the object o truly is morally good

In order to test this hypothesis, subjects were presented with a story about someone who has morally wrong desires, but seems to value something morally good. [Only condition 2 and 3 (motivation and guilt) are explicit].
“George lives in a culture in which most people are extremely racist. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of people of his own race at the expense of people of other races. Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he harms people of other races. And sometimes he ends up acting on these feelings and doing things that end up fostering racial equality. George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial equality and just act to advance the interests of his own race.” [Knobe, 2007, p.161]

Then subject had to decide whether they agree to:

*Despite his conscious beliefs, George actually values racial equality.*

In another version, “racism” was replaced by “racial equality”. Thus subjects were presented with a story about someone who has morally good desires, but seems to value something morally good. Then subjects had to decide whether they agree to:

*Despite his conscious beliefs, George actually values racial discrimination.*

Results showed that when the attitude is good, it is considered to be one of the agent's values; when the attitude is wrong, it is not considered as one of the agent's values. This result provides some tentative support for the view that “moral judgments actually do play a role in people’s concept of valuing” (162)

Many other explanations of the Knobe effect were proposed: conversational pragmatics, emotional biases or factors that have nothing to do with the nature of the concepts. But for Knobe, they are not “the result of any kind of ‘bias’ or ‘distortion.’ Rather, moral considerations truly do figure in a fundamental way in the issues people are trying to resolve when they grapple with folk-psychological question” (Knobe 2007, p. 57). A pervasive feature of folk-psychology is that people’s intuitions about intentional actions “do appear to be influenced in some way by moral considerations, but it does not seem possible to pick out any particular sort of moral judgment that we are aware of making and say: ‘It is this sort of moral judgment that is influencing people’s intuitions about intentional action’.” (2007b, 101). The best explanation so far, according to Knobe and collaborators, would be that people make non-conscious moral judgments, and use them to attribute intentionality: “we simply determine whether the behavior in question violates any of the norms in the very limited set we are considering. If it does, we classify it as a transgression” (2007b, p. 102)

In a study by Pizzaro et al, subject where presented with stories about (1) gay men French-kissing on the street and (2) “couples” French-kissing; subjects in condition (1) considered that the couple were intentionally promoting gay kissing (supposedly because people have non-conscious preferences about gay sex).

All this could suggest that calling our everyday interpretative practice a “theory” of a folk version of scientific practice may be misleading for many reasons (the same reasons that could explain why many were reluctant think of FP as a moral-normative tool instead of a theory). There is a common implicit assumption in philosophy and cognitive science, according to which folk psychology is folk science, and science is not about moral claim, but prediction/explanation. Thus FP is not and could not be primarily about morality. For example, if a scientist changes her
prediction according to moral preferences, her judgment is considered distorted, biased, not profoundly moral. But is FP a commonsense, intuitive science?

Science as we know it, however, is a recent cultural phenomenon. There is little reason to suppose that these practices “reflect anything fundamental about our innate cognitive endowments” (2007, p. 164). Hence if FP is compared to a theory, it does not mean that it functions like a scientific theory. As Daniel Hutto argues, FP is holistic and structured like a theory, but FP is more a “framework” than a theory. A system of beliefs can be theory-like: it posits unobservable entities to explain observable ones, it goes “beyond the surface”. But this is something that religions do, yet religions and scientific theory are profoundly different. Even if FP is a theory, it has not the same “sensitivity to empirical evidence” (Ibid., 166). Scientific method is, roughly about generating and testing hypotheses, and empirical evidence can be used to confirm or disprove the hypothesis. Theory-theorists argue that FP develops by falsifying mentalistic theories, nativists argue that there is no falsification, while others argue that it is an innate module parameterized during development (cf chapters 3 and 8). But sensitivity to empirical evidence is not specific only to scientific theories. Religious beliefs can be sensitive to empirical evidence, but no one claims that they are identical with scientific theories. The distinction lies mostly in what scientist do not or should not take into account. Scientists should consciously neglect moral considerations in adopting or revising a theory; but clearly these considerations are important in adopting or revising religious beliefs systems.

Another problem is that FP categories might not necessarily be best suited for scientific use (prediction-explanation); FP might create categories that serve other purposes. It may turn out “that the categories that prove most helpful in making moral judgments are completely orthogonal to the categories that prove most helpful in generating predictions and explanations.” (Ibid.168). Otherwise, we would need two systems: one for FP judgments (predictions), one for moral judgments (evaluations), plus rules for matching the two types of concepts. This is cognitively demanding, and unlikely. The “one-system that does both jobs adequately” solution does not work either: that would just show that FP is not like folk science. Science tries to develop precise, specific concepts tailored for prediction or evaluation, not general concept that does both job adequately. Hence folk theories are tools rather than predictive/explanatory framework (“theories”).
5.4 Folk-psychological accounts : a synthesis

We can now compare all the different conceptions of FP. Each account describes the basic entities that allegedly figure in the FP framework, the mechanism by which these entities are manipulated and the function of these operations.

<table>
<thead>
<tr>
<th>Entities</th>
<th>Psychological Mechanisms (operations)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard Account</strong></td>
<td>Beliefs &amp; desires</td>
<td>(variable)</td>
</tr>
<tr>
<td>TT</td>
<td>Beliefs &amp; desires</td>
<td>Theoretical inference</td>
</tr>
<tr>
<td>ST</td>
<td>Beliefs &amp; desires</td>
<td>Simulation</td>
</tr>
<tr>
<td>Gallager-Embodiment-Phenomenology</td>
<td>Sensorimotor capacities</td>
<td>Sensorimotor coordination</td>
</tr>
<tr>
<td>Hutto-Narrative</td>
<td>Beliefs &amp; desires</td>
<td>Narrative understanding and understanding narratives</td>
</tr>
<tr>
<td>Malle</td>
<td>Beliefs, desires, valuations, intentions, enabling factors, causal histories of reasons</td>
<td>Inferences</td>
</tr>
<tr>
<td>Knobe</td>
<td>Beliefs &amp; desires</td>
<td>Theoretical inference (but non-science-like)</td>
</tr>
<tr>
<td>Andrews-pluralism</td>
<td>Beliefs, desires, traits, emotions, moods, stereotypes, social norms, etc.</td>
<td>Inference</td>
</tr>
</tbody>
</table>

5.5 References

PART THREE: THE CONSTITUTION OF THE SOCIAL MIND

He who understands baboon would do more towards metaphysics than Locke. —Charles Darwin, (1838), Notebooks
6 Neurons that Mirror


Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
6.1 Explaining the social mind

The past sections presented philosophical and cognitive perspectives on the social mind: the philosophical problems related to interpretation, and the different accounts of interpretation. In this section (chapter 6 and after), we will be mostly concerned by the biological foundations of our social-cognitive competence: the evolutionary path, the adaptive function and the neural mechanism that underlie our interpretive faculties.

Explaining biological phenomena, as Tinbergen argued, might be done by four means. For any phenomenon (a trait, a faculty, an organ, a behavior, etc.), one may lay out its:

1. **Mechanistic** component: how the phenomenon is causally generated
2. **Development**, or ontogeny: how the phenomenon unfolds in the individual’s lifetime
3. **Evolution**, or phylogeny: how the trait evolved in the “Tree of Life”
4. **Adaptive function**: what the trait is for (if there is an adaptive function)

This chapter will focus on the neural mechanism of social cognition; however, since presenting these mechanisms would require an entire book, we will rather discuss a certain class of neural mechanisms that are particularly important for social cognition (and that generated many discussions in cognitive science and philosophy of mind): the so-called mirror neurons.
6.2 The brain

![Brain diagram](http://www.medem.com/MedLB/article_detaillb.cfm?article_ID=ZZZ0ZFP46JC&sub_cat=75)

Before discussing mirror neurons, a little introduction to neuroscience and its object, the brain, may be relevant (a good introduction can be found online at [http://thebrain.mcgill.ca](http://thebrain.mcgill.ca)).

The brain is the control center of the central nervous system. In human beings, it consumes 20-25% of the energy while it represents about 2% of the body mass (for a 150 pounds person). A 100 billion neurons are connected to between 1000 and 10,000 other neurons. This complex sensorimotor input/output device performs all psychological functions: sensation, pleasure, perception, emotions, reflex, motor control, motor planning, learning, representation, decision-making, imagination, reasoning, cognitive control, working memory, consciousness, etc. This mass of neurons shows a certain degree of specialization: brain areas tend to be involved in a certain type of function (e.g. the amygdala in fear, the prefrontal cortex in thought. However, as Anderson (2007) showed, the relationship between cognitive functions (e.g. perception, valuation, emotions, etc.) and neural structures is a many-to-many mapping:

1. **A cognitive function may recruit many brain areas.** For instance, social cognition recruits the anterior cingulate cortex, the amygdala, the prefrontal cortex, the medial temporal lobe, etc.
2. **A brain area may be involved in many cognitive functions.** The dorsolateral prefrontal cortex is involved in motor planning, organization, working memory, attention and many executive functions.

Thus the right attitude is not strictly localizationist (one-to-one mapping between function and structure) and not strictly holistic (functional equipotentiality in brain structures).
Neuroscience is the scientific study of the nervous system. Molecular, cellular, behavioral, and cognitive mechanisms are revealed through different means. Functional imaging technologies study degrees and locations of activation; single-cell recordings analyze the activity of individual cells, while lesion studies try to determine the function of certain brain areas by investigations of brain-impaired subjects. Computational neuroscience uses computer simulations to support or undermine hypotheses about brain mechanisms.

According to the classic DN (deductive-nomological) model, one explains an empirical phenomenon by showing how it “fits into a nomic nexus” (Hempel 1965:488). If, from the statement of a scientific law or a universal generalization (called also a covering law) and certain conditions one can logically deduce an event, this event is then regarded as being explained:

(1) L (scientific laws or universal generalization)
(2) C (condition)

(3) E (event)

But this model is now considered to be an inaccurate picture of scientific explanations, for different reasons. Which law, for instance, explain why cats land on four feet? Or the effects of a drug? Explaining any phenomena, in most sciences, means showing how it fits in the causal, not the nomic, nexus (Salmon, 1998). We want to know why molecule X causes inhibition of molecule Y uptake in the rat’s brain. The explanation of the inhibition of molecule Y does not cite a law, but a causal mechanism. A mechanism is “a structure performing a function in virtue of its component parts, component operations, and their organization” (Bechtel and Abrahamsen 2005). Mechanisms are described by their initial, intermediary and final conditions. Initial conditions specify the parameters that make the unfolding of the operations possible; intermediary conditions are causal chains (cycles, branching, network); and final conditions are the state reached by the mechanism (production, elimination, equilibrium). A mechanistic explanation of a phenomenon is therefore a specification of how the components of the system produce the phenomenon: it reconstructs the pattern of causal interactions that lead to certain results that need to be explained. For instance, explaining how the pancreas regulates blood glucose levels involves explaining how alpha cells of the islets of Langerhans lower blood sugar by producing insulin while the beta cells raise blood sugar by producing glucagons. The production of these hormones can also be subject to a mechanistic explanation.

Explanations in neuroscience tend to be mechanistic, multi-level and causal. That is, to explain a phenomenon, at any level of explanation, requires usually a mereological (part-whole) decomposition in entities and activities (Craver) or parts and operations (Bechtel). Mechanistic explanations are not (completely) reductive: higher-level phenomena are not reduced, but explained by lower-level phenomena. Explaining consists not, as the logical positivists put it, in a logical subsumption of a phenomena under a law, but in constructing a model of the systems and sub-systems that lead to the causal production of a phenomena. Mechanistic explanations, also contrast with traditional functionalism or computationalist accounts (or “two-levelism” see Lycan, 1987). Instead of positing only two levels of explanations (mind and brain, related by implementation), mechanistic explanations put emphasis on multiple levels of structure and
functions, not just two. In order to explain spatial memory, one needs to decompose this phenomenon in systems, sub-systems, sub-sub-systems, etc. A rough sketch (from Craver 2002) shows how four levels of descriptions articulate entities and activities.

1. Organism-behavioral: mice navigate through a maze to find food
2. Neuronal-computational: information processing, learning, connectivity of the hippocampus.
3. Synaptic-electric: neuron configurations (synaptic, dendritic) long-term potentiation
4. Molecular level-kinetic, attraction or phosphorisation by NMDA receptors

6.3 The Mirror Neurons System (MNS)

The **Mirror Neuron system (MNS)** is an action *execution/observation* matching system first discovered in the macaque *premotor cortex*, a brain area involved in motor planning. Visuomotor mirror neurons discharge both 1) when the monkey *executes* a motor act and 2) when it *observes* another individual performing the same or a similar motor act. The system does not react in presence of the object alone, the agent alone, the action without a target object, or the action performed by using a tool. Similar mirror phenomena have also been observed:

- audio-visual mirror neurons react to sounds that indicate an action (e.g. nut cracking)
- communicative mirror neurons: react to communicative facial expressions
- phonological mirroring: hearing engages tongue muscles involved in speaking
- sympathy - emotional mirroring: experiencing and perceiving someone experiencing disgust involves the same areas, i.e., same sector of the anterior insula (Wicker et al. 2003).

Thus it has been suggested that the processes of production and understanding of at least simple actions share the same neural machinery. We use our neural resources involved in doing A to understand that someone else is doing A; we internally and effortlessly reconstruct the motor program that led to A:

> Every time we are looking at someone performing an action, the same motor circuits that are recruited when we ourselves perform that action are concurrently activated.”

*(Gallese & Goldman, 495)*

Motor neurons are mirroring when they are activated by the perception of the action they typically perform.

Since the relationship between the mind and the brain is a hierarchical, multi-level nesting of systems and the relationship between cognitive function and brain function is a many-to-many mapping, we should understand mirroring as a “mid-level” mechanism: it is realized in many brain areas (in humans, in Brodmann areas 21, 40, 45) that also perform other functions, and the mirroring mechanisms participate in many higher-level functions.
6.4 Mirroring and interpreting

One of the primary functions of the MNS is production and understanding-through-reproduction of actions and emotional expressions. It is important to note that the MNS is sensible to actions, (movements that have a goal, performed with an intention) not mere movements. Mere visual perception does not provide information about the goal/intention behind the movement. Hence when one see a barking dog, imaging studies showed that it elicits only visual activity, not motor, because we cannot understand experientially (mirroring) the barking (see Gallese 2007 for a review). One might say that the category of actions is (almost) a natural kind for our brain. MNS systems are also necessary for imitation: reconstructing someone else’s action requires a fast reverse-engineering of motor plan, a function subserved by the MNS.

According to Gallese, the MNS is also crucial for language. We mirror at the “vehicle” level (hearing a sentence activates our tongue muscle and phonological motor plans) and at the “content” level: our understanding of many concepts is based on sensorimotor simulation possibility (e.g. according to Lakoff and Johsson, we need the schema of force or the topological representation of a container to be able to grasp modal and set-theoretical concepts).

Finally, according to many researchers interested in the MNS, it is involved in mindreading. The fact that autistic persons have a deficient MNS supports this interpretation. More precisely, the MNS has been seen by many (such as Goldman) as a vindication of simulation theory (ST). Remember that, according to ST, mindreading is a method, a skill, rather than a theory; we “impersonate the mental life of the target” (Gallese and Goldman 1998) rather than theorize about it. We allegedly use “as-if”, pretend, fake desires, process them through our decision-making system, and generate hypotheses about the beliefs and desires of the target agent. In prediction, a forward simulation infers action from beliefs/desires; in retrodiction a backward simulation infers the beliefs/desires from the action.
The MNS, according to Gallese and Goldman, might be involved in retrodiction. The MNS first encodes plans or intentions and then tags them as “my” or “your” action, depending on whether the system was internally (when I perform the action) or externally (when I observe the action) activated. When we observe an action, motor areas are activated but the movement is not completed. Externally activated plans are largely inhibited, or taken ‘off-line’ (but not always, as people who talk a long time together and get along well often adopt the same postures without noticing it). Subjects with prefrontal lesions (and, to a certain extent, babies) compulsively imitate gestures, suggesting that imitation is a default mode. Thus mirror neurons would be neutral (regarding the authorship of the action): they would simulate first the intentional action (or “naked intentions”), and then the agent. The process allegedly goes through 1) observing behavior 2) registering naked intentions 3) attributing agency (“who” mechanism: me vs. the other).

The MNS seems to support ST because it shows that action understanding is arrived at through an “inner imitation”. The MNS, say Gallese and Goldman “seems to be nature’s way of getting the observer into the same ‘mental shoes' as the target”. As Gallagher points out, however, intentions are always “already clothed in agency” (Gallager, 2007, p.70). We never have to decide who performs the action: we just see who is. An action is always experienced as someone's action. Thus the mirror resonance+who mechanisms underlie inter-subjective perception (as a personal-level phenomenon) rather than simulation. We experiment a direct perception, not a step-wise process of mental replication:

_The other person has an effect on us. The other elicits this [neural] activation. This is not a simulation, but a perceptual elicitation. It is not us (or our brain) doing it, but the other who does this to us (Gallager, 2007, p. 73)_

6.5 References


Tinbergen, N., On the aims and methods of ethology. Zeitschrift für Tierpsychologie, 1963. 20:

7 Social Primates and the Evolution of Social Cognition

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.


2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
7.1 Evolution and social cognition

Evolutionary biology, since Darwin, is based upon the principle of “descent with modification”: a blind selective process retain certain forms of file while discard others; the selection is based on the fitness of the inherited trait. Thus natural selection requires only i) variation in a population, ii) heritability of variable trait, and iii) selection of these variations.

Darwin draw inspiration from demographer Thomas Malthus, philosopher (and founder of economics) Adam Smith and geologist Charles Lyell. Matlhus showed that populations grow faster than their resources; Smith showed that population-level processes are not necessarily planned, but may emerge from individual interactions; and Lyelle showed that simple change, over long periods, can lead to important changes. The actual states of geological formations can help to explain the past. Thus in the “economy of nature”, as Darwin puts it, each form of live strive to survive and reproduce, and thus must adapt. Current living beings embody years and years of adaptation. The modern evolutionary synthesis was achieved by Dobzhansky and other biologist who integrate Darwinian theory with mendelian genetics (1937). Darwinism provides a general mechanism to explain the diversity and adaptivity of living beings—natural selection—and a guiding principle for organizing the mass of facts about them, the “tree of life”. The following figure represent ape phylogeny:

Thus our species did not "descend from apes": human and current apes share common ancestry Humans are hominins, primates, mammal, vertebrate, metazoarians and thus share certain trait with each of them.

Darwin's anticipated a revolution in psychology:

In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation.

In the Descent of Man, Darwin sketched an account of human faculty based on “Evolutionary social psychology”. Darwin argue (as many contemporary evolutionary anthropologists/psychologists) that intelligence and social life co-evolved.

A defining feature of social animals is that they face specific category of problems. Not only do the have to cope with ecological problems, (climate, food, environment, etc), they have social problem. To borrown Cheney & Seyfarth expression (2007, p.12): "most of the problems facing baboons can be expressed in two words: other baboons". Social life poses challenges such as (from Schaller et al., 2007):

- Alliance formation
- Status
- Self-protection
- Finding and maintaining mates
- Kin Care

Hence, given the assumption that natural selection operate also on mental faculty, social animals might be not only adapted to their environment, but also to life in society. Thus primate and human intelligence might be an adaptation to social life. Darwin suggest that our “social instincts” might have evolved and turned into complex reasoning capacities:

> Ultimately our moral sense or conscience becomes a highly complex sentiment-originating in the social instincts, largely guided by the approbation of our fellow-men, ruled by reason, self-interest, and in later times by deep religious feelings, and confirmed by instruction and habit. (Descent of Man, chap. 5)

About a century later, in "The social function of the intellect" Nicholas Humphrey draw a similar connection between intelligence an social life:

> “social primates are required by the very nature of the system they create and maintain to be calculating beings; they must be able to calculate the consequences of their own behaviour, to calculate the likely behaviour of others, to calculate the balance of advantage and loss -and all this in a context where the evidence on which their calculations arebased is ephemeral, ambiguous and liable to change, not least as a consequence of their own actions. In such a situation,"social skill" goes hand in hand with intellect, and here at last the intellectual faculties required are of the highestorder. The game of social plot and counter-plot cannot be played merely on the basis of accumulated knowledge, any more than can a game of chess”

Thus social-cognitive capacities are required. Clearly, our folk-psychology could be an adaptation to social life. But then, is it something peculiar to our linguistic nature, or does our close cousins also enjoy similar abilities? The question that haunted primatologists, since the famous Premack and Woodruff (1978) paper is "Does the chimpanzee have a theory of mind?"

### 7.2 Primate minds

Premack and Woodruff (1978) define the Theory of Mind (TOM) as a capacity to impute “mental states to himself and to others (...). A system of inferences of this kind is properly viewed
as a theory, first, because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behavior of other organisms (Premack and Woodruff 1978, p. 515). This characterization raises two kinds of problems.

Empirical:

Do chimpanzees predict and explain behavior?
If predict and explain behavior, do they use mentalistic concepts, or behavioral ones?
If they use mentalistic concepts, is it B-D?
(we know chimpanzee have certain concepts; it is not clear if they have mentalistic ones and if they use mentalistic concepts in social interactions).

Conceptual:

Are predictions and explanations the main functions of folk-psychology (FP)?
Do we always use mentalistic concepts to do so?
Are B-D the only mentalistic concepts?

How we define “theory of mind” (ToM) influence the question “Does X have it”. Is there evidence of a theory of mind in chimpanzees? Or are chimpanzees using only behavioral rules? The literature identify usually the possession of a ToM as manifest in imitation, self-recognition, social relationships, role-taking, deception and perspective taking, (Heyes, 1998). The challenge raised by what we might call “skeptical behaviorism”, is to see if one can reformulate behavioral evidence without mental terms, in which cases the observed behavior might not require a ToM. It many cases, it is not clear whether this capacity indicate a ToM or if another “behaviorist” explanation would be sufficient. For example, imitative behavior can be acquired by other means than mental-state inference. The sweet-potato washing observed in Japanese macaques is not necessarily an imitation, but maybe the induction of a behavior by similar stimulus (the macaque goes in the water, drop it, and conclude, like the other before, that it remove the sands), stimulus enhancement rather than imitation. Moreover, one ca imitate an action without attributing mental states. Self-recognition (the fact that apes, but not monkeys, touch their head in front of a mirror when they see a red spot that was painted while they were sleeping) maybe due to the fact that chimps touch their faces more often than monkeys, or because the experimenter introduces the mirror when the animal is completely awake (anesthetic artifact). Moreover, “mirror-guided body inspection” is not a sign of a ToM, but a sign that the ape have a “body concept” not a “self-concept”. Autistic children do not attribute belief but inspect their bodies. However, the problem with “skeptical behaviorism” is that even children who pass the false belief test have not ToM! The child could be using behavioral regularities about Sally “people look for things where they were”

As Kristin Andrews note, the problem might be also with the standard account of Human folk-psychology (Andrews 2005, 2007). According to this standard account, FP uses beliefs-desires for prediction/explanations. Thus it consists in 2 assumptions

(1) **Homogeneity**: one mechanism (mental state attribution) generates prediction and explanations

(2) **Narrowness of functions**: prediction and explanations are paradigmatic functions (cooperation, bonding, etc, are ignored in accounts of FP)
A problem with (1) is the plurality of mechanisms. We use several mechanisms to predict behavior, (and some of them are not mentalistic), such as:

Assuming people will act like us
Stereotypes
Generalizations about past behavior
Personality traits

Another problem is the inaccuracy of B-D predictions. When we focus on B-D, our predictions are less accurate and we overestimate the probability of actions when provide reasons. But we are generally good at predicting behavior in ordinary situations, so these predictions might not be primarily based on B-D. The problem with (2) is the plurality of functions. FP is used also for: evaluation and moral judgement (Knobe), story-telling (Hutto), etc. We use it for interaction, coordinating behaviors, detecting deception, bonding, etc. A Brandomian could also insist on justification: humans ask for reasons, and have to provide reasons. FP is the framework we use when we give and ask for reasons

Hence, the problem with the standard account is that prediction/explanation is one thing; attributing B-D is another; and FP is not exhausted by the set of all propositional predictions/explanations.

The question that should be asked is the: do chimpanzee use mentalistic concepts? It seems that they do, in certain practices that requires mentalistic concepts: tactical deception, punishment, bonding or coordination. Certain bands forage collectively and coordinate their effort by distributing functions between individuals. While they hunt, certain chimps can have 4 four roles: the driver, the chaser, the ambusher and the captor. This coordination seems to require a sort of FP for coordination, prediction, understanding, etc. Coordination is compatible with the mastering of FP concepts. We should not look for mentalistic concepts only in terms of predictions, but see if chimpanzees master any mentalistic concepts (for prediction or not, B-D or not) such as SEEING: If Al know that Bert see an apple, than Al attribute a mental state to Bert (Note that the question is not whether chimps can see, but whether they have the concept of SEEING). When subordinate chimps mate and feed when dominants cannot see them, this seems to indicate that they master the concept of seeing.
In an experiment, Hare et al (2001) used the following display:

![Diagram of subordinate and dominant chimps with occluders and food]

The subordinate sees the food, the dominant cannot. When food is occluded, the subordinate approach it; when it is not, he avoid it. The subordinate knows what the dominant knows, and have a concept of SEEING (“he does not see it, so I will eat it and won’t be punished”). Could it be explained by a non-mentalistic strategy? Maybe the subordinate use behavioral rules; “avoid food when the dominant is oriented toward it”, etc. However, the subordinate behave similarly even when the dominant is behind and opaque door: when the food is in the dominant’s visual field, the subordinate avoid it. Hence he does not read behavior here. The subordinate does not have a simple rule to decide whether he approach the food or not. He uses abstract generalizations about the visibility/ invisibility of the food (“when invisible for the dominant, go for it”)

According to Andrews, to understand the chimps representations of visibility as “that which the dominant may approach” we need mentalistic concepts. To know something about visibility is to know something about seeing. The chimp concept of seeing might be similar to the 3-years-old child: SEEING is not connected with BELIEVING, but clearly connected with an abstract mentalistic concepts. Thus chimps have some kind of FP because they see others as minded agents, and engage in certain FP practices

**References**


8 The Modularity of The Social Mind


Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
8.1 Modularity and the Mind

In his 1983 book *The Modularity of Mind*, Jerry Fodor put forth the idea that perception is in large part modular: instead of a general faculty, it is a constellation of many automatic and autonomous modules. Against “horizontal faculty psychology”—the idea of the mind is divided in faculty such as perception, memory, judgment, Fodor proposes a “vertical faculty psychology”: the mind is not divided in many modules, and each module has its own perceptual, mnemonic and information-processing capacity. Thus, for example, there is not a memory faculty, but many kinds of memory. A Fodorian module is

- Domain-specific (sensible to a specific type of inputs)
- Encapsulated (it has its own database)
- Fast
- Independent of voluntary control
- Relatively inaccessible to central systems not just in their informational encapsulation but also in being opaque to introspection
- The representations they compute are relatively shallow or non-conceptual
- Innately specified and have a fixed neural architecture
- Subject to specific breakdown

In a nutshell, modular systems are smart reflex: fast as a reflex, smart as cognition. The best example is the Muller-Lyer illusions (next figure): Even if we know that the 3 lines are the same length, we still see the illusion. According to Fodor, this is an example of a module, that automatically processes certain types of inputs, without any control over the process and without any possibility of a cognitive access to the module processing: we just “see” the output. Our beliefs that both lines are of equal length have no influence on our perception of the lines. Thus the database of visual perception is separated from our beliefs about cognitive illusions. According ot Fodor, the essence of modularity is **encapsulation**: a module performs its function autonomously, without beliefs or reasoning being able to influence the processing. Examples of modules include: language understanding, face recognition, object recognition. We recognize faces and objects, understand utterance without any conscious effort, without deciding it and without knowing how we do it (and this is why cognitive science and linguistics try to understand how our cognitive machinery performs so well in these task while we still don’t know exactly how to build machines that would be on par with us)

According to Fodor, central processes are not modular. Thus, for example, analogical thinking, scientific reasoning and beliefs revisions are not encapsulated and automatic, but require that we apply our higher-order deliberative capacities. Peripheral, perceptual modules send their outputs to the non-modular central system who process them.
8.2 Folk-psychology and modularity

Are belief-desire attributions modular? According to Fodor, there are no cognitive module. However, certain psychologists, cognitivists and philosophers proposed that there are cognitive modules, such as a module for the Theory of Mind (TOM). A TOM module, if it exists, performs fast, encapsulated, autonomous, social-psychological judgments.

For proponent of a TOM module, psychological knowledge and processing does not apply to other domains and can be selectively impaired and it is in large part innate. There are many reasons to posit a TOM module:

1. **Specialized domains**: Our knowledge of other persons, our folk-psychology, is distinct from our knowledge of animals (folk-biology) and objects (folk-physics). As Dennett forcefully argued, our application of the intentional stance is not reducible to the application of the other stances.

2. **Fast, automatic, mandatory**: We can understand see what someone wants without thinking explicitly about it

3. **Specific breakdown.** Autism suggests a specific impairment: autistic persons can still perform normally in non social task

Hence Baron-Cohen (1995) suggested that mindreading is subserved by a set of modules: an intentionality detector and an eye-direction detector that that fed into a shared attention mechanism which allow the operation of a theory of mind module. Autism is a shortcoming only in the two last mechanisms.
Another powerful justification for a TOM module comes from evolutionary psychology (EP), a research program that apply evolutionary biology to the understanding of the mind: our mind, just like our body, is a set of adaptations. The next section present the EP version of the argument for a modular TOM.

### 8.3 Evolutionary psychology and modularity

EP can be summarized by its four guiding principles:

- **Principle 1.** The brain is a physical system. It functions as a computer. Its circuits are designed to generate behavior that is appropriate to your environmental circumstances.

- **Principle 2.** Our neural circuits were designed by natural selection to solve problems that our ancestors faced during our species' evolutionary history.

- **Principle 3.** Consciousness is just the tip of the iceberg; most of what goes on in your mind is hidden from you. As a result, your conscious experience can mislead you into thinking that our circuitry is simpler that it really is. Most problems that you experience as easy to solve are very difficult to solve -- they require very complicated neural circuitry.

- **Principle 4.** Different neural circuits are specialized for solving different adaptive problems.

- **Principle 5.** Our modern skulls house a stone age mind.

Thus EP is a Darwinian, computational, massively modular cognitive psychology. Our mind is a collection of adapted module—a cognitive phenotype—each one solving an evolutionary problem. One can contrast EP with important rival theories:

- Against traditional cognitive science e.g. (Fodor, 1983), EP suggest that all the mind is modular. Cognitive science should be understood as a branch of biology. However, EP is not identical with sociobiology: organisms are adaptation executors, not fitness maximizers (Tooby & Cosmides, 2005, p. 14).

- Against what EP call the SSSM (Standard Social Science Model), the idea that our mind are primarily determined by cultural specificities, EP argue that our minds are not tabula rasa: cultural variability is limited by genetics.

- Against ABC (Associationist, Behaviorist, Connectionist) psychology, EP suggests that learning cannot start from zero, that computation must be modularized, that modules are innate domain-specific adaptations.

---

2 Evolutionary Psychology: A Primer, Leda Cosmides & John Tooby, online at: [http://www.psych.ucsb.edu/research/cep/primer.html](http://www.psych.ucsb.edu/research/cep/primer.html)
A paradigmatic example of adapted cognitive module discovered by evolutionary psychologists is the so-called Cheater Detection Module (CDM).

It was observed that performance logical task is higher in social context. Take this puzzle for example:

Here are four cards. Each of them has a letter on one side and a number on the other side. Two of these cards are shown with the letter side up, and two with the number side up.

E C 5 4

Indicate which of these cards you have to turn over in order to determine whether the following claim is true:

If a card has a vowel on one side, then it has an odd number on the other side.

Which one should be turned? If you answered E and 5, you are like most of the subjects: wrong. You should flip only E and 4:

- E card verifies the rule (if there is an even number, the rule was not followed)
- C card could not verify or falsify the rule: the rule does not apply since the antecedent is not satisfied
- 5 card could not verify or falsify the rule: The rule says that if there is a vowel, there will be an odd number; but if there is an odd number, the rule does not imply that there will be a vowel.
- 4 card could falsify the rule, because the rule implies the following IF no- [odd number] , THEN no-[vowel] . In flipping this card, we should see whether the rule was respected (no-vowel) or not (vowel)

Now look at another version of the question:

In its crackdown against drunk drivers, Massachusetts law enforcement officials are revoking liquor licenses left and right. You are a bouncer in a Boston bar, and you'll lose your job unless you enforce the following law:

“If a person is drinking beer, then he must be over 20 years old.”

The cards below have information about four people sitting at a table in your bar. Each card represents one person. One side of a card tells what a person is drinking and the other side of the card tells that person’s age. Indicate only those card(s) you definitely need to turn over to see if any of these people are breaking the law.
Instead of 25%, it is 75% of the subject who get it right this time. For Tooby and Cosmides, this is an evidence of adapted cognitive design for social exchange, or “social contract algorithms”: our reasoning is more sensible to contract or reciprocity violation than to the violation of logical rules. This violaters are cheaters: a cheater (or free-rider) is someone who enjoy the benefits without paying the cost; since our species lived in tribes for a large part of its existence, it is reasonable to expect our cognitive system to be able to distinguish easily between cooperators and non-cooperators, beneficent and free-riders. Since it is a recurrent adaptive problem, it also make sense to think that we have an adapted module.

Thus, according to EP, central reasoning system can be modular, domain-specific and encapsulated, just like the cheater detection module. Our capacity for mental state attribution is, according to Tooby and Cosmides, an adapted cognitive module.

8.4 The limit of the Modularity Thesis

One problem with this account, however, is that if our TOM is a central module, then we have no power over its output, and would not be able to resist any mental-states attribution, and unable to override an attribution. An agent with an automatic TOM would be too easy to manipulate. Gerrans (2002) suggest that evolutionary considerations argue rather for a domain-general capacity: we use a vast array of information in social understanding, and we can revise our beliefs about an attribution. These properties are the hallmark of a domain-general capacity.

Another argument for the modularity of TOM is that fiction seems to create a TOM illusion. Watching an actor whose character is in pain, the argument goes, create an illusion: we know he is not in pain, but cannot refrain form attributing pain. The problem here is that in fiction, we do not really believe what we see: we cannot literally believe (even with a small probability) that the actor is pain without being 1) epistemically and 2) morally deviant: otherwise, we would 1) believe what we know is false and 2) we would knowingly let someone suffering without doing anything. But clearly fictions are not “entrapment tools” (Currie & Sterelny, 2000, p. 151). Generally, our responses to fiction engage our affective and evaluative—but not doxastic—functions.

Other arguments points toward the Poverty of the Stimulus: we don’t see or perceive anything like a mental states, but all end-up with a theory of mind. Thus it might be an innate competence. The problem with this argument is that although propositional attitudes might be non-observable, the evidence licensing us to infer their presence is ubiquitous and universal. Even domain-general minds would converge toward the hypothesis that people have intentions.

An interesting suggestion by Currie & Sterelny (the “modest thesis”) is that social judgments are not produced by an encapsulated cognitive mechanism, but are influence by a second-order module, or a modularized layer between perception and belief fixation: an encapsulated perceptual mechanisms that tags modular output with social makers. Hence central (non-modular) processes can manipulate this information, without “buying it”: we see the actor as happy or angry, but do not literally believe it. The markes “push us in the direction of certain social beliefs, without itself being the mechanism that fixes that belief” (Currie & Sterelny, p. 155).

This mechanism—a perceptual sensitivity to certain social-like stimuli—can explain certain biases in reasoning with the same process:
• **Anthropomorphism**: We spontaneously attributing mental states to non-mental systems (animals, devices and illustrations) without believing it.

• **Intuitive trust/distrust, liking/disliking**: Our intuitive emotional or evaluative responses are not motivated by beliefs, but that can even lead to action (running from a terrifying individual)

The modest thesis present an evolutionary plausible solution. Social cognition is smart and fast, and must trade-off between these two constraints. A high-level system is computationally costly and many interpretation are rather simple and reliable (e.g. the meaning of smile), while a completely automatic, input-driven system is susceptible to manipulation (e.g. not all smiles are honest). One is smart but too slow, the other is fast but dumb. A second-order module is smart and fast:

“Tagging the perception with an intentional marker serves to push the belief fixation system in the direction of one conclusion rather than another. Without strong countervailing pressure from other beliefs, we shall conclude quickly and efficiently that this is a friend. But the intentional marker always leaves room for such countervailing pressure, and for the consequent rejection of the friendliness hypothesis.” (Currie & Sterelny, 156)

This model is also compatible with facts about autism: autistics subject can still form beliefs; they just can’t have certain beliefs about other agents. Their impairment might be rather perceptual than cognitive (“mindblind” rather than “mind-illiterate”). Autistic persons can recognize human faces, (lower-level) but less apt than controls to identify their emotions (social tagging). Similarly, Gerrans suggest that FP is a non-modular synchronization of modules.

**References**


9 Social Intelligence

Eskimo Seal Hunting (date unknown) Peter Winkworth Collection of Canadia,
http://commons.wikimedia.org/wiki/Image:Eskimo_Seal_Hunting.jpg

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.

2008 Benoit Hardy-Vallée
www.hardyvallee.net/POSC
9.1 Human Intelligence: puzzle, chess or collective hunting?

Needless to say, our species is one of a kind. We have culture, Internet, writing, language, satellites in space, mathematics, universities, science, arts, etc. We live in every possible climate. We have a large brain able to store a vast amount of information and to solve many problems. How can we explain human uniqueness and its evolution from the first hominid to our world?

Three accounts of human intelligence will be discussed here:

1. **The ecological intelligence hypothesis.** Humans are smart because they are adapted to the complexity of their environment.

2. **The social intelligence hypothesis:** Humans are smart because they are adapted to life in society: complex sociality generates complex minds, which increases social complexity which increase mental complexity, and so on.

3. **The socio-ecological intelligence hypothesis:** Humans are smart because they collectively adapted to their environment; social and ecological fused.

Each hypothesis can be illustrated by a different activity: a jigsaw puzzle (humans facing a complex world), chess (humans facing other humans) and collective hunting (human facing a complex world together).

9.2 Ecological and Social Intelligence

Proponents of the Ecological Intelligence Hypothesis (ECH) depict our intelligence as an adapted capacity to process complex factual information about the world. We are good at pattern recognition and color vision because the challenge of foraging (finding and extracting food) require a lot of learning, memory and inference. Foraging for seasonal fruits requires botanical knowledge; nut-cracking, requires technical knowledge, etc. Moreover, as Richard Potts (1996) argued, the African genesis of the hominid clade took place in conditions of increasing climatic instability, and our species, in its geographic and ecological expansion, faced changing, challenging environments.

While the ECH has been a prominent thesis, many psychologist, primatologist and antropologits favored a Social Intelligence Hypothesis (SIH). Primates are “social experts” or social problem solvers, rather than problem-solvers tout court. Social life comes with a lot of challenges: the formation and monitoring of a network of relationships, cooperation, communication, social understanding, reciprocity, dominances, membership, deception, manipulation, etc. The complexity of social life promote the complexity of individual cognition, and vice-versa: there, according to proponents of th SIH, a co-evolution between social and cognitive complexity. That Explain the “extraordinary transformation”, in about 4 Millions years, from our ancestors who “lived in the world as they found it” (Sterelny, 2007) to our modern societies and our distinctive intelligence. Primates intelligence would be an adaptive response to the complexity of social environment where primates evolved.
As Sterelny shows, there are at least two reasons why this hypothesis is plausible. First, the complexity of our social world: cooperation inter and intra-generations and the division of labor. Our societies are vertically complex (in terms of hierarchy) and horizontally complex: there is a lot of differentiation at each level (individuals, families, clans, tribes, communities, nations, cultures) in expertise, economic role. Second, there is a feedback loop between culture and cognition: social intelligence increase the complexity of the social world, which select for greater social intelligence. If we can never detect lies, treasons, manipulations, chances are that our species would be extinct. Proponent of the Machiavellian intelligence hypothesis (a version of the SIH) suggest that social chess generates social complexity and vice-versa. Proponents of the Machiavellian Intelligence Hypothesis (MIH, e.g., Byrne & Whiten, 1988, 1997) argue that the advanced cognitive processes of primates are primarily adaptations to the special complexities of their social lives, rather than to nonsocial environmental problems such as finding food” (MITECS)

Or, as Sterelny puts it, the evolution of hominin cognition depends on features of the environment that hominins have created themselves (2007). Our “Machiavellian” intelligence evolve through an arms race of manipulation, counter-measures, counter-counter measures, etc. that lead to greater social intelligence. A standard way of illustrating this is the Prisoner’s dilemma.

In this game, the police hold, in separate cells, two individuals accused of robbing a bank. The suspects (let’s call them Bob and Alice) are unable to communicate with each other. The police offer them the following options: confess or remain silent. If one confesses –implicating his or her partner—and the other remain silent, the former goes free while the other gets a 10 years sentence. If they both confess, they will serve a 5 years sentence. If they both remain silent, the sentence will be reduced to 2 years. Assuming that Bob and Alice have common knowledge of each other’s rationality and the rules of the game, they should confess. Even if they would be better-off in remaining silent, this choice is suboptimal: they would risk a 10-years sentence if the other does not remain silent. Experimental game theory showed that, contrarily to orthodox game theory, subjects cooperate massively in prisoner’s dilemma, even in one-shot games (Ledyard, 1995; Sally, 1995).

The PD is an example of a collective action problem, such as people rushing out of a building in case of fire. All should calm, but then one could get out faster; knowing this, however, everybody rushes out. The PD applies to any situation where everybody would profit from cooperation, but that would make the temptation of defection too attractive. When the PD is indefinitely iterated, however, cooperation can be rational. Axelrod’s (1984) computer simulations showed that cooperation can be rational if one follow a cooperative reciprocal strategy: cooperate first, cooperate with cooperators, do not cooperate with defectors

According to the MIH, since humans cooperation is both risky and valuable (there is always a risk of being cheated), our evolved psychology dispose use to monitor our interactions with other and to engage in “tit-for-tat” altruism: we cooperate with those who cooperate, and do not cooperate with those who doesn’t. We are endowed with calculative-cooperative disposition, where reciprocity is sustained by self-interest and mutual scrutiny.
9.3 Socio-ecological Intelligence

One problem with the Machiavellian account is that calculative cooperation requires “cooperative score-keeping”, memorization of past interactions, etc. As Sterelny suggest, the Machiavellian theory is a better description of our contemporary, anonymous mass societies where we have to evaluate our engagement with strangers. Our ultrasociality, however, evolved in non-anonymous social worlds. Human cooperation, he argues, is not calculative, but a default option evolved for hunting and defence coalitions (Sterelny, 2007).

The socio-ecological intelligence present an alternative model for cooperation, based on 1) Strong Reciprocity, 2) Symbolically marked groups and 3) Norms of cooperation

**Strong Reciprocity**

In direct reciprocity (tit-for-tat) model, cooperators compute (consciously or not) the worth of cooperation. However, many findings in experimental economics suggest that Humans are default rather than calculating reciprocators, such as the ultimatum game and public good games.

In the **ultimatum game**, a ‘proposer’ (Alice) makes an offer to a ‘responder’ (Bob) that can either accept or refuse the offer. The offer is a split of an amount of money. If Bob accepts, he keeps the offered amount while Alice keeps the difference. If Bob rejects it, however, both players get nothing. According to a standard interpretation of the game, rational agents must behave as follows: Alice should offer the smallest amount possible, in order to keep as much money as possible, and Bob should accept any proposed amount, because a small amount should be better than nothing. Thus if there is $10 to split, Alice should offer $1 and keep $9, while Bob should accept the split. The ultimatum has been studied in many contexts where different parameters of the game were modified: culture, age, sex, the amount of money, the degree of anonymity, the length of the game, and so on (Oosterbeek et al., 2004; Samuelson, 2005). The results show a robust tendency: the game-theoretic strategy is rarely played, because people tend to make ‘fair’ offers. While proposers offer about 50% of the amount, responders tend to accept these offers and reject most of the unfair offers (less than 20%). They are thus ready to lose money in order to punish unfairness. In a **public good game** players may contribute an initial endowment to a common pool where it get bonified or free-ride by contributing nothing and receiving their share of the bonified common pool. While self-interest recommend free-riding, subject participate massively and punish free-riders, even in one-shot games (Ledyard, 1995; Sally, 1995).

Fehr and his collaborator coined the term ‘**altruistic punishment**’ to describe situations where people punish defectors, at a cost to themselves, through a non-utility-maximizing action, even in anonymous one-shot interactions or when they did not interact with the free-rider. Altruistic punishment, as research on public good games experiments showed, enforces cooperation: free ridings is less common and contribution to public good rises in groups where punishment is possible (Fehr & Gachter, 2000, 2002; Smirnov, 2007, p. 490). Punishment promotes cooperation by making it **rational** since defection has a cost and **predictable** since conditional cooperators can expect others to cooperate (Shinada & Yamagishi, 2007).

Hence we are Strong reciprocator rather that direct reciprocator: we cooperate by default, cooperate with cooperators, do not cooperate with non-cooperators, and punish cheaters.
Symbolically marked groups & norms of cooperation

Clans, tribes, associations, etc. all mark their affiliation through language, custom, values, cultural artifacts, tattoo, symbols, decorations of tools, etc. Different groups, from street gang to business associations, have different symbolic practices that allow them to recognize each other as belonging to the same “tribe”. Instead of memorizing and computing before cooperating, we cooperate more easily with those who identify themselves as “ours” (with all the problem that tribal instincts may generates, of course).

But deeper than symbols, norms is one of the most important cooperation-enforcing device. Norms make cooperation rational, and does not require “cold computation” of benefits: “rule-following” is easier than calculation. Moreover, cooperation is facilitated, and defection more obvious, in a society with explicit cooperative norms, such as food sharing between hunters. In a cooperative world, it is easy to cooperate, and easy to detect free-riders.

Rather than remembering past interactions and calculating the “cooperation score” of someone humans decide who to cooperate with. Cooperation, Sterelny argues, should be construed as a collective action, not a sequential exchange of favours. For most of their evolutionary history, humans lived in small foraging bands. Security and food, among others, were secured through n-person public goods games and not dyadic interactions (as in reciprocity models). Hunters and vigils do not compute each of their partner cooperative score: they assume that everybody will cooperate and comply with the norms of cooperation.

According to Socio-Ecological Intelligence Hypothesis (SEIH), humans are not purely ecological or social problem solver, but both. Our species is one of social foragers: there is no clear distinction between ecological and social intelligence. The cooperative, technology-dependent foraging typical in our species induce need for social skills: learning, communication, imitation. Collective and Big Game hunting requires a fusion of social and ecological skills.

Foragers must have information of the animal and their technology (ecological skills) and must also be social experts: mutual knowledge, communication and coordination, and division of labour are necessary for collective success. The acquisition of foraging skills also necessitate social skills: social learning, language, imitation, teaching, etc. These cognitive tools for culture were selected for their use in the acquisition and deployment of ecological skills. In learning foraging skills, the social and the ecological merge in the economic domain:

*Individuals did not face the ecological filters on their environment alone, but with others, and with the technology, information and misinformation that their social world provides. Ecological and social complexity became fused, as the ecological problem of extracting resources as individuals from a world we did not make became the economic problem of extracting resources collectively from and in a human world.*

What is the function of a TOM (Theory of Mind) in this Socio-ecological model? Principally, it is part of the toolkit for cooperation. With a TOM, we can use other minds as sources of information about the world. More importantly, TOM make teaching possible more efficient. One need to understanding the skills for teaching it, and the teacher must understand...
what the student knows in order to monitor its progress. In facilitating communication and the transmission of skills, TOM is an adaptation both to the ecological and the social environment.

References


• Whiten, A. Machiavellian Intelligence Hypothesis, entry in MIT Encyclopedia of the Cognitive Sciences (MITECS)

Bibliography on Philosophy and Social Cognition

Philosophy – Classics

Philosophy of mind


Gordon, R., 1986, Folk Psychology as Simulation, Mind and Language 1, 158-171;


Hobson, R. (2007). We Share, Therefore We Think. In D. Hutto & M. Ratcliffe (Eds.), Folk Psychology Re-Assessed (pp. 41-61). Dordrecht: Springer.


See other references at http://consc.net/mindpapers/7.1

**Animal Social Cognition**


**Evolution and Social Cognition**


**Economics and Social Cognition**


**Mirror Neurons**


Moral Cognition and Moral Neuroscience


**Social Neuroscience**


**Social-Affective Neuroscience**


**Social Cognition and development**