Can "radical" simulation theories explain psychological concept acquisition ?

Joëlle Proust

Published in ***Simulation and knowledge of action***
J. Dokic & J. Proust (eds.), Amsterdam : John Benjamins, 2002,201-228

Can "Radical" Simulation Theories Explain Psychological Concept Acquisition?

Joëlle PROUST
CREA, Ecole Polytechnique, Paris

The question of understanding the behaviour of other agents in a psychological or mental way divides into two sub-questions:

(1) How can a human child above 3 and a half or an adult attribute certain *kinds of psychological states* rather than others to people, i.e. a state of belief, rather than of desire, hope, fear, etc?

(2) How can such an interpreter find out correctly what is the *content* of the others' psychological states, i.e. a belief that P, a desire that Q, etc?

Both theory-theory (TT) and simulation theory (ST) aim at answering these questions, but each actually only offers one part of the explanation. TT explains the capacity of using psychological concepts as a result of an innate, specialised module or as a result of a general process of theory building. But id does not explain how psychological concepts may adequately be applied to specific contents.[1] ST explains how an interpreter can grasp the specific content of the other's

[1] See Heal (1995), 37, (1996), 75.

psychological state, while leaving in the dark how a simulator can understanding the various psychological attitudes that other subjects entertain towards that content.

In this paper, I will concentrate on the response offered to the latter problem by Robert Gordon, who espouses a simulationist view on how understanding others is achieved. This author indeed pushes to its limits the possibilities of *radical simulationism,* i.e. the view that mentalisation is a process requiring initially no possession of psychological concepts. It is interesting to see exactly what this theory can accomplish, and if it finally fails to explain mentalisation, *why* it does.

Radical simulationism : main claims

The basic claims of radical simulationism will be better articulated against the background of a weaker version of ST, such as the one defended by Alvin Goldman (Goldman, 1989, 1992, 2001). According to Goldman, simulating another agent consists in taking her point of view on the world, and in producing "offline" the various responses (decisions, emotional states) that one would produce in such a situation if it was actually present. The basic feature of simulation, considered as a representational process, is that it is *isomorphic* to the target-process, which it is meant to mimic (Goldman, 1992, 108).

According to Goldman, simulation is thus a three-step process. 1) Projecting into someone else's situation. 2) Appreciating introspectively what it is like to be in that situation. 3) Applying whatever psychological concepts available to categorise the feelings, desires and beliefs and other evaluations made in the second step, and attribute them

to the simulatee. Simulating, in such a theory, requires a prior expertise in mental state attribution. You must be able to recognise, say, a specific emotion or a belief in yourself to be able to ascribe it to another person. You must have categorised your own psychological states to recognize them as applying to other agents.

Robert Gordon diverges from Alvin Goldman's view on simulation on three major counts.

1) The first point of divergence has to do with the kind of *projection* needed to simulate others.Is the target of a projection a mind, a situation, an idiom? It will be useful here to contrast Gordon's view on projection with Quine's and Goldman's. Projecting oneself, in Quine's approach, implies that the simulator should imagine what it is like to be in another's state of mind. By this Quine understands that the simulator reconstruct the *system of sentences held true* by the simulatee. This process consists in translating the other's idiom into the simulator's, and in deriving the sentences that hold in the idiom modified in the appropriate way, according to a rationality principle:

> We project ourselves into what, from his remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned (Quine, 1960, 92).

Alvin Goldman would agree that projection has to do with imagining being in another's *state of mind*, but he would deny that projecting always involves an exercise in translation. He explicitly allows simulation to account for non-propositional mental states like pains and itches (1995, 84). Goldman maintains that simulating requires

introspecting what oneself would feel and believe if one was in such a situation, and ascribing similar contents to the simulatee.

Robert Gordon does not see projection as allowing any kind of introspection, nor would he indeed accept without qualification the view that in simulation, an agent generally projects into somebody else's *state of mind.* On the other hand, Gordon does not take Quine's view on projection as a translation from an idiom to another. Projecting consists rather in recentering oneself into another's location and concerns through what Gordon calls an "egocentric shift". A "personal transformation" into the simulatee occurs, rather than a transfer into somebody else's mental states (Gordon, 1995, 56). Simulation occurs through an imaginative transformation into the other, which allows the simulator to see the situation that the simulatee faces as *she* faces it. One might here object that it is not quite clear how such a shift may occur if not through some mental transfer. A plausible reconstruction might be the following. A simulator projects herself into a situation-type, i.e. in the situation that typically causes the mental state of the simulatee and makes it the type of mental state it is, rather than into the mental state itself (although projecting into a situation-type leads *de facto* to having some consonant set of beliefs and desires).

2) The second major point of divergence has to do with *mastery of psychological concepts being or not a prerequisite* for simulating others, and more generally, with the relationship between simulating and believing. In Alvin Goldman's view, a simulator needs to be able to identify conceptually the kind of the mental state that she introspects in the pretend mode, in order to assign it to the simulated person. Simulating presupposes further an ability to frame the

proper kind of situation to be simulated, an ability that involves a considerable amount of knowledge of what others might believe and want about the situation.

Gordon recognises that simulating situations different from one's own sometimes require a mastery of psychological concepts. He maintains however that there is a large class of simple cases in which no such mastery is required because no motivational or epistemic adjustment need to be done; furthermore, the capacity to simulate in those simpler cases might explain how psychological concepts are finally acquired. Early forms of simulation thus seem to him to require no mastery of psychological concepts, but rather, combined with "ascent routines", to allow a first crude way of categorising intentional states. The core idea of an ascent routine is the following. Phrases such as "I believe", "I want", etc. are first uttered purely expressively, in the absence of psychological concept possession: beliefs, wants, desires are expressed through a linguistic form borrowed from ordinary adult conversation, without the child having any particular concept of what the corresponding propositional attitude is.

Ascent routine was first described by Gareth Evans, in his *Varieties of Reference*, (1982) in a chapter devoted to *self-identification.* "In making a self-ascription of belief, he writes, one's eyes are, so to speak, or occasionally literally directed outward - upon the world" (p. 225). To know whether I believe P, "I have to put into operation whatever procedure I have for answering the question whether P". There is a semantic ascent in this process, because from P, i.e. a fact which I take as holding in the world, I can derive a fact about myself, which I assert through the words " I believe that P".

Now as both Gareth Evans and Robert Gordon insist, using this procedure does not amount to a "full understanding of the content of the judgement "I believe that P". For Evans, a subject who performs the ascent routine may still lack the mastery of the generality condition, which applies to psychological concepts such as "x believes that P". In other words, she may fail to understand that the content she believes, namely P, can also be believed by other people. She may also be unable to grasp the fact that a belief is a claim for truth, which can be supported or not by the right kind of evidence. What the ascent routine does achieve, is a way of bypassing the notion that information about the self has an inner source, to be accessed through introspection (Evans, 1982, 230).

For ascent routine to work in the case of primitive simulations, one needs only accept that a subject who is able to "label" a belief or a desire in a real situation is also able to label those attitudes in pretence. Robert Gordon acknowledges that this early kind of simulation does not deliver a genuine, comprehending ascription of mental states, but he nevertheless claims that such simulatory practice may foster "greater conceptual understanding" (1996, 17). He suggests further that a process of embedding ascent routines in simulations "gives sense to the idea of a mental location". When simulating Peter, John projects himself in Peter's situation, and evaluates it from Peter's point of view. Projection amounts to a process of accepting propositions believed by Peter given a situation S. John thus gets a notion of "something's being a fact to Peter". Projection is normally followed by a deduction of those consequences of the accepted propositions relevant to Peter in S, and by an ascription to Peter of a belief in the propositions thus derived. By embedding ascent routines in simulation, Gordon means

to refer to the iteration of the question "is it a fact that P?" in various perspectival situations. The simulator is able to store the various responses to the same question using various individual-based situations: is it a fact that P for Peter? Is it a fact that P for Mary? Etc. What John finally learns is that some facts are acknowledged by most individuals, some others by few. What he will be unable to learn in this way, Gordon notices, is that there are facts *without* a mental location, i.e. facts considered by nobody, and about which the question of truth or falsity as a matter of fact does not arise.[2]

3) Reliability — The question that is raised by the role of practical simulation in mentalisation is of what makes self- and other- ascription of beliefs and desires generally *reliable*. A process is reliable if it tends to produce true beliefs. As Gordon observes (Gordon, 1993, 45), reliability in self-ascription of mental states may be taken to stem from a decision procedure - provided by appropriate concept application in theory-theory, and by introspection and categorisation in Goldman's approach. Gordon favours a view in which reliability does not rely on a decision procedure, but rather on a non-inferential, non-conceptual process: verbal labelling as described in the ascent routine provides reliability of self-ascription without any concept being actually called for. In his words, "reliability does not need theoretical knowledge".

[2] This may be a problem for a full grasp of the concept of belief, but it may not be an objection to a theory of folk psychological concepts. It seems plausible that people tend to deny spontaneously that a proposition that nobody believes or even considers as possibly true indeed refers to a fact.

Objections to radical simulation theory

A first well-known problem with the view that a simulator only needs to use the information contained in some perspectival situation is that some mastery of epistemic notions must be present for a subject to select, during deduction, those premises that are accepted by the simulatee. In other words, the notion of a situation relevant for belief ascription cannot be defined in purely extensional or subject-neutral terms. A simulator must have some notion of which premises will be used in a simulated piece of practical reasoning. For example, if we take the tube test,[3] which end will Peter choose? The unconnected opening just under the spot where a ball is launched, or the farther but connected opening? To predict adequately where Peter believes that the ball will go, I must include in my premises the proposition that Peter accepts the view that the ball has a causally restricted dynamics. If Peter accepts instead that a shorter distance between launching and picking up is a relevant element in the problem, my deduction will take another course. It is hard to see how such an appreciation of what premises are relevant can be effected without presupposing the very notion to be explained, i.e. the notion that a subject *believes* those propositions which he accepts and acts upon.

A second difficulty concerns the view on psychological concept acquisition taken to result from a diversified simulatory practice. Gordon takes "embedding" of simulations to consist in a rather straightforward inquiry of whether the fact P is holding from the various perspectives of

[3] See Hood (1995).

individual thinkers. The simple structure of "embedding" which Gordon uses, however, falls short of providing notion rich enough to prepare the understanding of a *mental* location. In his view, decentering one's perspective on a situation only requires changing places, in the literal sense of varying the spatial distance and perceptual access to objects, and keeping track of the various views on enduring objects and states of affairs. Spatial memory allows storing various points of view on one and the same situation, along with the specific properties associated with each one of them. Given an observer endowed with objectivity (i.e. capable of grasping the fact that perceived objects are stable individuals existing even when they are not perceived), a subject is in this analysis able to grasp perspectivalness. When I was there, the object looked round. Now I am here, and the object looks oval.

Let us grant that simulating a situation does not involve taking the point of view of any specific subject, but only presuppose access to the informational content which one gets from occupying various possible locations. It may seem that a simple spatial variation between what is seen from here, from there, etc., offers the required contrast between situational properties, and prepares the acquisition of the concept of belief.

But this seeming is illusory. It is a product of the silent use of psychological concepts in the description of spatial changes. The verb "seeing" carries the burden of conveying a psychological information on top of the information collected about the scene. More strictly described, spatial change appears no longer as sufficient to provide the sense of a mental location. Clearly, as Strawson (1959) showed, changing one's view on the world constitutes a necessary condition for representing stable entities that may exist

unperceived. But Gordon's claim is that perspectival simulation is a sufficient condition for understanding mental perspective. Data from primatology support rather the idea that there is a cognitive gap between understanding perspective and understanding belief.[4] Non-human animals are quite good at predicting, in a simulatory way, which rival is likely to grasp some food, or to reach some prey, without having the least disposition to attribute to others the notion of "P being *a fact for X*" (or *X* believing P). Chimpanzees, for example, are certainly acting as if they had such a notion. But careful experimental work by primatologists (Povinelli, 1996) shows that it is not the case. Although they can use some psychological information in the sense that they can follow the gaze of conspecifics, and use that information to detect for themselves what there is to be seen, they are unable to understand that the other acquired some perceptual knowledge. Non human primates are unable to tell apart an informed partner from a naive one; they cannot infer directly knowing from seeing. Although they may learn how to hide some behavioural signs of their inner states, they seem to do so on the basis of associations between behavioural cues. Not: "If I shout, he will know". But:" If I shout, he will bite". Autistic children seem also to have trouble to infer from a perceiving subject that this subject knows perceptually something (Baron-Cohen & Goodheart, 1994). On the other hand, they have no trouble finding out what will be on the picture, when the objects in front of the camera are being changed, a task which plausibly involves simulating prior perspectives (Zaitchik, 1990, Leslie & Thaiss, 1992). Mastery of perspectival information does not amount to mastery of epistemic perspective.

[4] For a detailed discussion, see Proust (2001a).

These findings tend to show that *extracting* psychological information - simulation being one way of doing so - may guide behaviour without allowing the agent to understand that the information used is indeed mental. For all the animal knows, so to speak, the information used is behavioural. It has to do with the types of things others can do in a particular situation, not with what the types of things they can believe and want to do. Therefore changing perspective through simulation does not offer *per se* any clue about psychological capacity.

Let us come finally to Gordon's view on reliability. The notion of reliability that Gordon has in mind does not amount to a capacity of providing *rational grounds* for what one says or does. It has to do with the fact that the agent, while being or simulating being in a situation, will *normally* identify the right content of his/her belief or desire. After some linguistic training, children are able to utter "I want a banana" when and only when they want a banana. Doing so does not amount to a "genuine self-ascription of desire", Gordon acknowledges, because they may still lack a concept of desire (as well as a concept of themselves). This analysis illustrates the fact, stated above, that simulation theory is better equipped for explaining how particular psychological contents can be understood than for categorising the attitude which the simulated subject has towards that content. True, a child manages to convey that she wants a banana, rather than an orange; doing so, she does not need to know that there is something she wants, rather than hopes, regrets, believes, etc. But this simply reflects a fact about training; a trained animal reliably does what it was trained for doing. Why should this tell us anything about psychological concept acquisition?

Therefore the reliability of the mechanism for *uttering* sentences that *reflect* mental states (but actually do not *refer* to them) should be contrasted with the reliability of a mechanism for self-attributing mental contents. Younger children may communicate what they want reliably, while utterly lacking a capacity for self-ascribing correctly psychological state types. Is the first kind of reliability a step towards acquiring a reliable mechanism of the second type? We shall struggle with this problem at the end of the chapter.

Simulation reconsidered

Gordon's radical simulationism stems from three important intuitions. The first is that simulation may be performed in a context-driven and agent-neutral way. The second is that simulation is a very basic, multi-purpose psychological process, that does not necessarily involve psychological concepts. The third is that simulation is harnessed to a practical reasoning system. From these three basic intuitions, the challenge consists in showing that simulation may indeed lead to psychological concept acquisition, or at least may provide an essential element in mentalisation.

A) Simulation as a context-driven process

Any kind of action involves representing a context, or a situation, in which the corresponding type of action normally develops. When an agent aims at producing a certain result by acting in a certain way, he/she has to represent what can be called "the canonical context" for that action. Such a canonical context includes a certain type of movement or

active involvement, as well as the relevant objects and properties (or the most salient) for reaching the given goal. If an agent's goal is to get food, for example, she will have to represent the canonical context most appropriate for that goal, which involves both perceptual cues and propositional knowledge (including concepts and inferential relations between them): is this cake on the kitchen table available? Is it fully cooked? Should I go instead to the bakery? Etc. Memories of prior successful actions determine what constituents will be included in a given canonical context. A context may be represented, in a few cases, in an essentially non-conceptual way (for example, when it comes to a particular step to be executed in dancing with a given rhythm and tempo). It may also, in other cases, involve mostly conceptual knowledge (such as, for example, the kind of events to expect when taking an exam).

Most contexts involve both kinds of contents, conceptual as well as non-conceptual. Such an "action space"[5] organised in the perspective of an acting subject is the fundamental representational structure activated and exploited in simulation. As simulation theorists have insisted, a capacity to simulate necessarily depends on the experience acquired by a subject: a given representation of a context includes elements that may not play a role in practical reasoning, or may be supplemented, replaced by other items, etc. When planning an action, an agent will need to simulate various contexts for reaching her goal, and various courses of action within one context. In the simulative phase, the agent needs to articulate specific perceptual expectations with conceptual-instrumental reasoning. A young child may at first have trouble identifying the essential ingredients constituting a

[5] Trevarthen (1999).

specific canonical context. Observing repeatedly the actions of other agents, listening to stories, engaging in pretend play, are the various means which help a child engage actively in context-typical simulations in ways that are more and more specific and appropriate.

Empirical evidence tends to show that human agents represent actions, motor sequences and instrumental relations between means and goals (i.e. canonical contexts) in a way that is "agent-focussed" and "subject-neutral". By these expressions is meant that an observed action is represented in the perspective of an agent, rather than in the perspective of the particular person who observes the action (who may be the acting self, but also somebody else). Granted that a canonical context of action articulates a goal, means-ends relations and specific perceptual cues, it seems particularly important that the specific orientation of the observer's body does not present a systematic obstacle to understanding the actions of others. Brain imaging studies (Decety *et al.*, 1997) support the view that, in both cases, the dynamic representation of a context - its simulation - are performed under the agent's point of view. *Even when someone else's action is observed, the observer activates for himself the dynamical representation of the corresponding action in the neural structure that normally allows execution*.

This does not mean that the action is fully performed by the observer; nor that understanding another subject's action presupposes introspection of the same action performed by self; this only means that the "active" coding of an action is involved in performing as well as in observing it. A subject who has to identify whether a presented hand is a right or a left hand will move mentally her own hand - at a subpersonal level. A subject watching someone take an object with two fingers will activate a similar pattern in her *premotor* area.

Seeing an action seems to constitutively involve a disposition to act in a similar way.[6] Although observing an action rarely evokes a conscious disposition to act, there are circumstances - e.g. observing a soccer game or watching a thriller movie-, in which the observer clearly feels immersed in the action he watches, and actively projects into it in a way that is consciously accessible. He catches himself running and jumping.

B) Simulation as a psychological process

From a psychological point of view, simulating consists in running "off line" stored information. Such an information may be perceptual and either refer to facts retrieved in an imagistic (analogic) mode, or to particular aspects of those facts (one can simulate having an experience of blue, of heat, of pain, of drinking coffee etc.). It may be pragmatic or praxic,[7] with again the same contrast between whole contexts and individual properties: one can simulate oneself as doing this and that, or as moving in a particular way (in a more or less richly imagistic way). It can also be propositional, and simulation can be conducted on the basis of compositionality of senses -, as is the case in planning. The feature common to these simulatory episodes is that their contents are represented as outcomes of a subject's concrete abilities or doings rather than in an abstractly inferential way. In this respect, simulation has much to do with episodic memory (Tulving, 1983). For example, while simulating a future action (an activity that essentially involves concepts and

[6] See Daprati *et al.* (1997), Jeannerod (1999), Proust (2000).

[7] Pragmatic information concerns the kind of action performed, whereas praxic information concerns the kind of movement performed when acting.

propositions), the planning agent considers the consequences of a selected context in a practical way (how the situation will then appear to her), rather than as detached predications (made so to speak "from nowhere").

Simulating is thus a basic kind of dynamic memory process used in various areas of procedural knowledge, that cannot be divorced from the conceptual capacity used by a subject to categorise and infer external events and properties. The difference between semantic memory representation and simulation is simply that the former proceeds in a purely conceptual way, while the latter represents contexts in a mixed way. It uses both *concepts* to categorise a given context and *qualitative, first-person experience* to characterise the feelings and emotions relevant for the overall evaluation.

A second important point is that in each of its specific domains of application, simulation can also be used in many different ways according to what we might at least *prima facie* characterise as the reasoning capacity available to the subject. Through simulation, as we already saw, various kinds of experienced canonical contexts with their associated inferences are extracted from memory and combined. In some systems it will be mainly used to revive past embodied situations; in others, it will be used in reasoning about future or counterfactual contexts.[8] As several theory-of-mind-theorists observed,[9] simulating can be of any help to a mentaliser only if it allows deploying the various kinds of

[8] The role of counterfactual reasoning in theory of mind acquisition is also stressed in Riggs *et al.* (1998) and in Donald Peterson's contribution in this volume.

[9] Leslie & German (1995), 129 ; Perner (1996).

inferences relevant in the simulated situation. The scope of reasoning in an individual thinker might here be constrained by the executive dispositions of the individual: those necessary for maintaining distinctive representations of one and the same context separate, and for exploiting the inferential relations across them.

C) Practical reasoning, counterfactual reasoning and simulation

There are two ways of simulating a counterfactual situation that need to be contrasted and explored further. Overlooking this difference may explain in part why Robert Gordon's radical simulationism fails to state the conditions that lead from simulation to mentalisation. To articulate this distinction, we need to explain what the *frame of reference* of a simulation is.[10] A frame of reference consists in the modality in which an agent deploys her representation of a given context. A frame of reference can be the real world (in that case the subject represents past or present, actually experienced context tokens or types). It can be a potential world: a future context token is represented as relevantly analogous to a past one. It can also be a non-real state of affairs, i.e. counterfactual or strongly counterfactual: in the latter case only, the subject represents explicitly the situation as part of a possible world distinct from the real world.

The capacity of mastering the various modal dimensions of thinking in distinctive frames of reference has interesting links with the capacity to *reason conceptually*. The generality principle, in virtue of which mastering a concept implies the ability to entertain "indefinitely many thoughts" in which this

[10] For a more detailed presentation, see Proust (2001).

concept appears (Evans, 1982, 105) would seem to require from any concept user that he should already be capable of counterfactual reasoning. This type of reasoning seems to be indeed a major way of exploring in thought the various combinations allowed by conceptual components. This requirement would have the unfortunate consequence to restrict concept use to those thinkers that are able to represent possible states of affairs in exotic worlds. For example, is it needed, to master the concept of a "unicorn", to understand sentences such as "had they been real, unicorns would have superseded horses in number"? A weaker interpretation seems compatible with the generality principle, a principle which, according to Evans himself, is "an ideal to which our actual system of thoughts only approximately conforms" (ibid. 105). In this weaker sense, mastering a concept would imply an ability to entertain thoughts involving that concept inside at least one frame of reference. Accepting this view on concept mastery would lead to various extensions of "generality" according to the set of objects and contexts included in the world (or sets of worlds) represented in thought. This kind of distinction would account for the fact that a child may reason about unicorns inside the story which presents them as having some properties, while being unable to consider them "outside the story", as fictional entities with no direct causal impact on the real world.

What is true of concept use also applies to episodes of simulation. Not only because an episode of simulation often involves concepts, as we saw above. But also because simulation, being a species of representational thought, also possesses a frame of reference for its own sake, independently of the concepts it contains. It may be the case that imaging a shade of blue, for example, presupposes having actually seen the same shade (just as applying a

concept to a new object presupposes having the concept). But nothing seems to restrict imaging to retrieving prior percepts. Imaging may also combine imagistic contents in new ways (as presumably, does a composer or a painter when creating a new piece). An artist, who can exploit these differences and their implications as part of what his/her piece suggests to an observer or a listener, can also, in such a case, consider several frames of reference in parallel. It may be supposed that the informed observer/listener will be able in turn to trace these various hints to contrastive approaches when exposed to the work.

Robert Gordon was right to stress that simulating was a basic way of performing practical reasoning.[11] But he apparently failed to observe that the crucial step for simulating mentally consists in the joint ability to represent *arbitrary* states of affairs *and* to reason *across frames of reference*. Simulating familiar, absent contexts is the main focus of pretend games. Children, around 2 years of age, exercise their practical reasoning ability in putting together counterfactuals and deriving relevant consequences in a world close to the real one (Leslie, 1987). It should be noted, however, that the frame of reference of story telling or of pretend games is restricted to the isolated situation considered. There is only one world being referred to when a boy pretends being an Indian warrior, and the reasoning involves considering exclusively the properties that hold inside that world. Thus a playing child's way of simulating uses the same kind of capacity as when remembering familiar contexts (both kinds of memory involve representing only one context in a world). The only, but important difference is that the child engaged in pretend-play calls this

[11]See also Heal (1996), 78.

representation actively, rather than in a passive associative mode. In pretend-play, some properties of the props are actively disengaged while others are retained (a banana does not hold as a fruit, but as an object shaped as a telephone). Fiction also has the property of helping the child to project himself in a situation including a number of arbitrary properties which will be used as a basis for inferences - those being constitutive of the understanding of the story. There is no transparency of psychological attitudes at this stage in phylogeny or in human development: mammals, primates and young children exercise pretence without being able to fully understand what pretending involves. They know how to pretend without necessarily knowing that pretending conceptually imply believing.[12]

Exercising this imaginative ability in story understanding or in pretending however does not amount yet to a full understanding of pretending as a mental state. For such an understanding would require, in the view I defend, the further ability *to move within and without the simulated situation*.

In a primary stage of simulation, as found in pretend-play, no explicit representation of the world as it truly is, nor of the contrastive claims of truth made in the real world and in the pretended world, need be involved. This interpretation of pretend-play differs from a view in which make-belief presupposes that, in every episode of pretending, there is a set of beliefs incompatible with the pretended situation, which the pretender must explicitly rule out while she pretends. Evans (1982) for example analyses make-believe as involving both engaging in pretence *and* suppressing

[12] For a similar view, see Heal (1995), 40.

disbelief. In that analysis, simulating that * F has G* (when as a matter of fact, F does not have G) implies that one believes that "F does not have G" and *that one silences that belief*. As a consequence of this view, thinking counterfactually in pretence involves representing two worlds: the world of the game, and the world as it independently is. It also requires having a set of rules (a rule of incorporation and a recursive rule) establishing which truths from the real world still hold within the pretence.

In the present analysis, simulating that *F has G* does not *initially* presuppose that "F does not have G" and that the subject knows it. Weak counterfactual thought, therefore, consists in simulating a situation in which F does not have G, without representing explicitly whether doing so violates the rule for correctly representing F. In such a case, a canonical context for acting is presented as satisfied, using partial cues and ignoring those cues that do not match. The key element in pretending seems to consist in highlighting the relevance of actions, gestures and social meaning as determining a shared canonical context, to the detriment of various perceptual contents. The canonical gestures and their targets determine saliences in the perceptual scene. Pretending that one is in a car might be obtained by imaging a car and producing some auditory signal for the car's engine. A child pretending to drive a car does not need to actively combat the belief that he is not actually sitting in a car. The pretence develops from the simple fact of imaging the dynamical properties associated to driving.

This analysis is consonant with developmental evidence. In pretend play, as well as in understanding narratives, children have initial difficulty at maintaining the scripts separate: they tend to take the products of their imagination

as real, in particular when they are afraid by them.[13] They may also infer that some properties in the world have been changed as a result of the events in the fiction. Children finally learn that the *world of the game* and the *world in which the game occurs* are different, and license separate kinds of deductions. It is of course not clear how they learn this (to know how to respond to this would amount to having accounted for theory of mind acquisition). It may be through continuous practice, with the help of a more and more effective working memory. Alternatively, it may be through the help of a specific innate module coming to fruition around 3.

What is suggested is that there is more in strong counterfactual make-believing than "simple" pretence, and that the difference can be understood in terms of the frames of reference involved. This ability has been described, in Recanati's apparatus, by the distinction between *exercising* a simulation and *exploiting it.* Although we need not accept the theory of projection/retrojection defended by Recanati, the contrast drawn by Recanati is indeed close to the present view on strong counterfactual reasoning.[14]

---

[13] Harris *et al.*, (1991).

[14] According to Recanati, *exercising* a simulation consists in accepting a set of premises as holding in a particular imagined situation, and drawing from them some relevant consequences inside the same imagined situation. Accepting the premises does not mean that the simulator must take them as *true.* He only *pretends* them to be true, or *imagines*, or *supposes* them to be true. Deriving further properties from the supposed ones does no more belong to pure imagination. Such a conditional reasoning may be taken as involving typical relations within the real world or a subset of them.

Exploiting a simulation allows considering the same situation both from the alien point of view in which we project ourselves (through exercising simulation) and from our own point of view. Exploiting simulation allows

If we now examine the structure of the specific type of simulation involved in psychological attribution, it appears that the concept of a frame of reference is essential in understanding exactly what it is to embed an ascent routine in another. What Gordon suggested is that, by successively engaging in a number of projections, one becomes able to grasp the notion of a mental location. The problem we identified above is that nothing constrains the simulator to entertain the notion of a causal role for belief. She might as well reason, as chimps probably do in their own modes, that agents have varying dispositions correlated with the behavioural and other physical cues available in the environment.

This is no longer possible in the case of the kind of embedded counterfactual reasoning that we are presently describing. In the exercising mode, the simulator must first project into a situation in order to draw various *"internal"* inferences from the properties that it contains. In the exploiting mode, the simulator must examine the external consequences of the projected situation for the real world (or for some other counterfactual world). For example, in a task

---

the simulator to consider the situation simulated not only as a set of meaningful relations between properties, but also as having certain relations with the actual world. For example, understanding Mallarmé's poems, requires both exercising and exploiting simulation, i.e. requires a parallel consideration of the situation depicted, and what belongs to the writing conventions. This is not a situation peculiar to literature. Understanding mentally another person requires both seeing things as she does and retrodicting the causes or predicting the real effects of the simulated situation in connection to the world as we see it. While exercising a simulation involves a projection within a situation, exploiting a simulation involves an embedding of the former situation within a real world perspective, a process which Recanati calls *reflection* (1999, 101).

of unexpected transfer, the child must first represent the counterfactual situation CS according to which the object did not change its location (from O to O'). She must then draw the consequences of this situation for an action relevant in CS (fetch the object in O), and finally infer from this the consequences for the real world: the simulated agent will go to O, even though the object is now in O'.

What may be an obstacle to understanding the role of exploitation in false belief tasks is that spatial properties are easily traceable from one world to the next: in the allocentric mode, locations can be determined absolutely. The unexpected transfer task relies on the ability to determine absolute locations, and memorise which locations have been salient in a past, now counterfactual situation. In such a task, the extra step from examining a counterfactual world to drawing consequences in the real world may well appear to be a useless sophistication.

Cognitive lying offers a finer-grained way of exploring how a subject exploits a simulation. The interest of this kind of ability is that it cannot rely on spatial memory alone. Let us suppose that Jean murdered his wife in their Parisian flat, but claims that he was away from Paris at the time of the murder. In order to make his lie accepted as a truth, the simulator must engage himself (through a narration or some kind of staging) into a counterfactual situation such that the desired properties follow: he was away watching an exhibit in London. But he also must compare the world of his pretence with the real world, in order to revise, among the many properties derivable in both worlds that are incompatible, those that can be found not to hold in the real world. Inter alia, all the signs of his presence in Paris must be erased, the Channel ferry or other public transportation systems must have been working, the exhibit must be on in

London that day, etc. To be successful, the liar must both exercise simulation of the fictional world he projects into, and exploit the real consequences of his fictional world, through maintaining a maximal coherence across both worlds. He will have to use the kinds of rules which Evans spelled out for make-believing: new facts can be incorporated in the lie as long as they do not contradict any statement (falsely) presented as true.

Exploiting a simulation thus first requires establishing in thought two types of representations: one of some pretended situation and the other of a model of the real world as it appears to the simulator. Then the simulator has to reason across these models. He has to evaluate the non-actual situation on the background of the real world, to derive which facts are or can be made compatible and which facts are incompatible (ignoring the property that forms the core of the lie). This step may be seen as a truly embedded simulation, or as two embedded uses of the same simulation. The question can now be raised again of how this more complex kind of simulation allows an agent to acquire psychological concepts.

Complex simulation and psychological concept acquisition

In the present approach, a practical understanding of the concept of belief is supposed to be constituted by the capacity of evaluating the content of one simulation through the content of another. A subject who exploits simulation in the sense indicated has to draw inferences at least at two levels, allowing her to determine what is wrong in the simulation1 compared to a simulation2. Is the concept of belief, with its inferential structure and its associated truth-value, practically grasped as an outcome of the kind of embedded

counterfactual reasoning described above? The practical know-how involved consists in exploiting the simulated perspective for one's own sake; i.e. deriving consequences inside a simulation that, properly redescribed, are relevant in the real world. This view is in agreement with a prediction of evolutionary psychology that psychological concept acquisition is part of the arms race driving selection of new capacities in phylogeny. Using what one understands as being conspecifics'views on things is of primary significance in order (i) to take advantage of what a foe knows or fails to know by manipulating him into improper actions; or (ii) to help a friend or a relative to acquire knowledge or revise his beliefs in order to act rationally.

There are two obvious questions to raise at this point. One is: why should the choice of the relevant assumptions in simulating some foreign perspective be any less circular in the present suggestion than in Gordon's one? The second is: how is a simulator able to grasp from his embedded simulation a notion not only of belief, but also of all the various psychological attitudes that can be entertained?

1) Simulation theory and vicious circularity

I will suggest that what appears as a vicious circularity might result from a perspective mistake. What could be termed "simple" simulation occurs only early in development (or in other primate species). Simple simulation is simulation conceived as a process reflecting only knowledge of physical facts and, for humans, linguistic expressions for occurrent belief states that are not fully understood as mental states. In other words, simple simulation is the kind of simulation of others that is exercised in the absence of psychological concepts.

Now it is clear that, at some point in development, psychological concepts and theories do help a simulator to adjust her simulations to the specific cases in which the projected situation departs on certain assumptions from the corresponding closest real world situation. On the other hand, it was claimed above that the really important step in simulation occurs when the situation to be simulated is not of a familiar kind, i.e. does not coincide with the memory of a past event experienced by the simulator. The difficult question is therefore the following: Can *this* kind of simulation - using both a counterfactual situation and another world of reference to which the former is compared - get off the ground without using psychological concepts?

Fortunately, there is a class of cases showing that it indeed is the case. The reason is obvious: the kind of simulation based on counterfactual reasoning that is necessary for mentalising is also needed in cases having nothing to do with psychological attribution. While planning a complex action, for example, one must be able to introduce new objects with a set of consequences (exercising simulation), and predict what the consequences on the planned course of action will be in a specific real-world context. For example, when planning to clean a well, one must consider the possibility that some toxic gas be present at the bottom and kill the cleaning human agent. This type of simulation involves various inferences *external* to the situation considered, through a process of exploitation of the simulation. It also needs a revision of the default assumptions that hold when cleaning an ordinary object. Such a revision of the situation does not need to appeal to any psychological knowledge.

If pure counterfactual simulation is available at the level of the physical world and exercised routinely in planning

action, it may safely be concluded that this type of reasoning does not involve as a precondition a mastery of psychological concepts. In the case in which simulation is used to understand others and predict their behaviour through psychological attribution, one could suggest that the type of counterfactual reasoning involved is of the same kind, except that situations are indexed to definite individuals "whose situation it is". One could object here that "a fact to x" already involves an understanding of a mental access to x. But remember that a radical simulationist will export, so to speak, his own immediate epistemic relations to facts, appropriately restricted to specific contexts. The objection can therefore be countered by the routine ascent: I deal with P1, Peter deals with P2, and we form representations corresponding to P1 and P2. The notion of "dealing" as used by the simulator in fact draws on his own belief/desire structure, not necessarily on his belief /desire concepts. For the rival representations in question dot not need to be *represented as representations*, contrary to what theory-theorists would claim.[15] They only need to be accessed demonstratively, as "this situation" versus "that situation".

Saying this does not commit the simulation theorist to the bold view that all there is to be known about belief can be gathered through this practical ability of embedded reasoning. The idea is that the practical reasoning ability is the foundation for understanding others, not that acquiring this ability through reasoning on various counterfactual contexts is *sufficient* to reliably predict others *in all cases*. Knowing the kinds of facts that are typically encountered by others is often not enough to simulate them; you also need to know what they specifically believe, desire, or fear to

[15] In particular, Perner (1991).

represent the situation as they see it and to predict accurately what they will do. It is thus plausible that the set of assumptions to be used in deriving the facts that hold as a consequence of some situations cannot in many cases be correctly established when the simulator does not fully master the concept of belief. Simulation theorists do not need however to deny this obvious fact. What they might suggest is that a child starts applying strong counterfactual simulation in a less-than-reliable way. An *imperfect* capacity for simulating others may be enough to get the practical know-how required for belief understanding started. In a situation such as the false belief task, for example, a simulating child only has to represent dynamically the goal and dispositions for action attached to the context as known, and those attached to the context before the transfer occurred.

But the preceding parallel of simulating others with counterfactual planning may give rise to a more serious objection. The latter focusses on the claim that reasoning across frames of reference, - exploiting the consequences of one frame in another -, should *provide* the simulator with the practical knowledge of what it means to have a *false belief*. The objection[16] goes as follows. Let us imagine that A is able to reason across frames, and that he compares how things look to B with how things are. In such a scheme A can attribute a false belief to B, but this is a far cry from attributing false beliefs *to himself*. Considered as concepts, truth and falsity apply however *in a general way* ; such a generality might well fail to hold in the case of the kind simulation presently explored; even though simulating counterfactual situations involves a capacity to evaluate a

[16] I thank Jerôme Dokic for articulating this objection.

possible world as not realised, there is no guarantee that a simulator will thereby be able to evaluate his/her own simulations and predictions as subject to error. Recognizing that one's *own* beliefs can be false, the objector concludes, involves more than reasoning across frames of reference; it involves understanding that beliefs are propositional attitudes, i.e that their contents can *constitutively* be true or false. Therefore our view on simulation plus counterfactual reasoning is not better off than Gordon's radical simulationism. Both views are indeed circular.

Before attemping to answer this powerful objection, let us first remind what is at stake. The present paper, again, does not aim at defending the view that simulation might provide in and by itself all the necessary means for understanding other minds in all circumstances. What it does claim, however, is that a practical way of grasping and possibly putting to use another subject's representation of the world consists in *exploiting* counterfactual reasoning; i.e. drawing the consequences of a situation as simulated in the simulator's own world. As we will see below, this ability could be the underpinning of mental concept acquisition. Clearly, if it is shown that counterfactual reasoning of this kind can only be performed by a subject who is already mastering the concept of truth, then our claim will fall prey to circularity. Several authors indeed seem to accept the view that a subject only grasps the concept of truth when she possesses a theory of mind (Carruthers, 1996, Papineau, 2000) .

Now one way of defeating the objection is to show that there is a conceptual difference between evaluating practically one's own embedded simulation and explicitly applying to it a metarepresentational concept of truth. Accepting this conceptual difference does not lead us to expect however that

is should be observable in empirically distinct predictive behaviors or rational strategies. Even though only a full-blown theory of mind can provide a subject with the fully general and explicit concept of truth, a simulator endowed with the kind of counterfactual reasoning described would indeed reach a coextensive practical knowledge. For truth is something that a simulator desires even though he does not master the concept of truth. A simulator wants her simulation to deliver predictions from which she can actually take advantage. As Papineau puts it, "If you act on true beliefs, you will generally get the results you want, but not if you act on false beliefs" (Papineau, 2000, 201). In the case of simulation, this implicit, or practical desire for truth, needs to apply to one's own simulation. If you try and figure out what another subject is up to, you must be prepared to *revise* your simulation. A search for alternative explanations involves an implicit recognition that a given episode of simulation got things wrong. The distinction that the individual simulator needs to master at this point is of a successful/unsuccesful prediction[17].

This argument from the very structure of revision in simulation can be completed by phylogenetic considerations on the origin of mentalizing capacities. It is generally accepted that the selective pressure on more and more sophisticated mentalizers is based on the need to detect cheaters, liars, misinformed and malevolent communicators. Thus a reflexive dimension should be immediately salient in

[17] It will be noted that the appearance/reality distinction, as studied by Flavell et al. and many other developmental psychologists, does not coincide with the distinction between two stages in acquiring the concept of truth. Our notion of counterfactual simulation allows a simulator to grasp practically this contrast as one of differing simulated points of view.

counterfactual reasoning applied to others : knowing that one can be simulated and taken advantage of, by being offered a distorted view of reality is a crucial step in such reasoning. In order to understand that someone tries to lie to you, you must simulate yourself listening to a false indication, and acting on its basis in a way serving the other's ends. Thus a successful simulator is one that anticipates that he could be manipulated into misrepresenting things ; but again, this anticipation does not have to be explicitly understood in terms of truth or falsity, nor even in terms of representation; it is enough if the subject is able to imagine (or recall) the situation A where P holds, and contrast with situation B where P does not hold, with their contrasted affordances for the self or for others.
Let us summarize this point. Exercising simulation does not require a concept of simulation or representation, any more than believing requires having a concept of belief. Having the sense of a successful or unsuccessful simulation is the key to revising, but no concept of truth or of a true belief needs to be mastered. Finally the simulation of being possibly deceived is the essential reason why counterfactual simulation is being developed; this gives us a presumption, if not an argument yet, that a simulator is able to look at her own simulations as fallible.

The general solution of the problem of psychological concept application is that reasoning mentally involves a capacity to use perspectival information in strongly counterfactual cases. The latter ability is linked to executive properties of working memory, (maintaining active two rival representations for the same context) and not to psychological concept possession per se. Simulationists do not need to deny that this kind of reasoning should be considerably enhanced through linguistic expression. Symbol

use will obviously help a child to build the relevant metarepresentations, to reason from them and to communicate rationally about them. This observation leads us to our second question.

2) Is a simulator able to acquire the various psychological concepts through simulation?

We saw earlier that Robert Gordon admits that simulation does not deliver a genuine, comprehending ascription of mental states, while nevertheless claiming that simulation may end up with "greater conceptual understanding". (1996, 17) We tried above to clarify in which way secondary simulation involves a specific kind of counterfactual reasoning. The question that needs to be raised at this point is to understand whether this *know-how* develops into a *know-that;* is a simulator able to grasp the general structure of reasoning involved in mentalising, and to identify conceptually the various roles which propositional attitudes play in a subject's reasoning? How does a simulator manage to distinguish the various psychological concepts (believing, desiring, intending) that may be predicated of the various simulated contents?

As Heal remarks, Robert Gordon's view that simulating just implies putting oneself in a state similar to the target fails to provide the kind of psychological knowledge that a theory of mind of any variety should provide:

> "If getting into states similar to those of others is to be of any use in psychological understanding, it must not merely occur; it must also be recognised for what it is" (Heal, 1995, 44).

How then might we construct a simulator's understanding, say, that John *believes* that P, without positing

circularly that the content of the simulator's pretence was [believing that P]?

A plausible suggestion for bridging the gap between the simulatory episode and psychological concept use might consist in invoking a procedure accepted in other fields of concept acquisition, i.e. demonstrative ostension. Heal presents this strategy as a general solution for identifying both the content of a simulation episode and the kind of simulation instantiated in this episode. What subject A simulates is that subject B entertains *this content.* A also actually simulates that P *in this way* (*believing*) rather than *in that way* (*desiring* or *fearing* that P). This solution however presents some problems of its own, (on which we will not expand here) having to do with the fact that simulating as an experience (referred to demonstratively) is not itself a belief experience (but rather a pretend-belief experience). Therefore pointing to a simulating episode will not allow pinning down a belief episode, but at best a pretend-belief one. At best, however: for when engaging under the pretend mode in a given situation, a simulator will normally have a variety of attitudes, from pretend-belief to pretend-desire and pretend-fear, not to mention various pretend-emotions. An additional difficulty is that the notion of a belief (or pretend-belief) *experience* is not clearly independent from the corresponding conceptual thought that one has a belief, or a pretend-belief. Thus the disintrication of contents and propositional attitudes linked to a given situation seems to be made no easier through the demonstrative procedure alone.

A more promising way of disintricating attitudes may consist in relying both on simulation as a reasoning procedure and on social learning. We saw earlier that the ascent routine, as usually understood, is only a way to parrot, rather than to express a judgement self-ascribing a belief.

What makes it so is that the use of *believing* that is relevant in the routine ascent expresses only a product of a primary simulation (or simple simulation) of a specific context. A "secondary" simulator however (able to perform strong simulation) transforms her prior understanding of the word "I believe", when developing her reasoning abilities. The question becomes one of articulating the know-how involved in reasoning *counterfactually* and the conceptual knowledge about belief and other psychological concepts.

A distinction made by Andrew Woodfield (Woodfield, 1996, 1997) will help understand how such an articulation may be effected in the course of development. We will first present his view, presented as the author did in a way consonant with theory-theory, and will adjust it later in terms consonant with simulation theory. The psychological concepts that are constitutive of a theory of mind, such as believing, desiring, and intending, should be seen as accessible public concepts, i.e. essentially communal norms. According to Woodfield, concepts in general are of this kind. They regulate human reasoning and communication in a way that does not presuppose their being explicitly grasped under a theoretical formulation:

> "The player subscribes to the norms in advance of fully knowing what they are and without knowing in detail how to apply them, just as a club-member subscribes to the rules of the club without having properly read the rule-book" (Woodfield, 1996, 91).

On the other hand, in his view, an individual thinker has a particular set of inter-related beliefs and know-how, however vague and inchoate, about any specific domain (folk-psychology as well as folk-physics, folk-biology, etc.). Her beliefs about belief, desire, and related concepts are a

result of her particular experience, and thus differ from any other thinker's in some ways. These beliefs are theoretical in the sense that they go beyond observation. These individual theories, which Woodfield calls "conceptions", are "evanescent" personal constructions, in the sense that they are permanently revised and thus highly unstable, particularly so in the course of development between 2 and 6. A child may thus both have a common concept of belief in the sense that she understands sentences of a public language containing words expressing it, while possessing a different, idiosyncratic personal conception of belief (in which, for example, whatever true beliefs the self has are equally accessible to other selves). The important two points are first that, even in such a case, the child can be credited with *having the common concept of belief,* rather than with using some private or distorted concept; and second, that her conception of belief may become *more appropriate*, i.e. more *reliable*, when her individual conception gradually adjusts to the common norm.

Woodfield's approach can be accommodated within a simulationist framework. The core idea is that psychological concepts form a framework of socially accepted concepts and norms, to which every language user refers when communicating about psychological facts and events. In the course of development, a child needs two different types of information to apply these concepts in a reliable way to new sets of facts and events.

She must first have access to the relevant kind of linguistic exchange: "the ascent routine" is one of many standard ways of mastering a mentalistic language. Others of

the same ilk are offered by parental guidance[18] in helping their children recognising emotions and various mental attitudes. The social character of these exposures to conceptual terms provides the content of concepts, although these concepts may only at first be used in very restricted kinds of inferences. In the present view, linguistic exchange provides a "proto-theory" for folk-psychology rather than a full-blown inferential structure. In the case of the mental, a crucial factor for building a psychological theory consists in what we called strongly counterfactual simulation.

The second kind of knowledge a mentalising child needs is thus the set of reasoning, representational and executive capacities that allow a thinker to use the social concepts in a fully general way. This kind of knowledge is simulatory and procedural. A child must know how to represent a type of situation across worlds and use this representation to make accurate predictions on how these worlds interconnect. Thanks to the second kind of learning, a child becomes able to become a creative mentaliser: she now has her own inferential capacities allowing her concept of belief to operate in new situations.[19]

---

[18] Cf. the role of mimicry in typing mental states is hypothesised in Gergely & Watson, (1998). See also Proust (2001).

[19] Although we cannot develop this point here, reliability of simulation procedures for understanding others' mental states may be articulated in terms of the specific conditions for simulating: that the processes of thinking across subjects (simulator and simulatee) are the same or sufficiently similar ; that the context simulated is the same or sufficiently similar; that the knowledge of the context is the same or sufficiently similar; and finally, that the simulator has the reasoning and executive memory capacities allowing him to reason across worlds in the way described above. (See Heal, 1995, 40).

As suggested above, both kinds of knowledge might actually be interdependent, in the sense of contributing to each other's development. Sentences retrieved from memory help simulate types of situations, and contribute to develop efficient execution. On the other hand, reasoning helps a child distinguish in which relevant aspects two linguistic descriptions of a situation type differ. Saying this does not amount to acknowledging that simulation does not have a structuring role in allowing a child to mentalise.

Conclusion

Let us summarise the solution offered above to the problem of psychological concept acquisition. This solution seems in its core to be in the spirit of Gordon's radical simulationism. The basic ingredients include again simulatory practice and ascent routine. Our expanded theory, however, contrasts two kinds of simulations, and emphasizes the structuring role of counterfactual reasoning in embedding simulations and deriving facts stable across them. Furthermore, it completes ascent routines with social learning of folk-theoretical words.

When exposed to a public language, a child is provided with sentences relating desire, belief, and intention expressions. The various terms involved in a theory of mind are first grasped in a shallow and idiosyncratic way as a result of the child's developping mastery of public language. As we saw, such a linguistic mastery of social concept-terms does not amount yet to possessing a theory.

Independently from her linguistic ability in using ascent routines and echoing adult's talk about belief and desire, a child relies on her developing reasoning and executive

memory capacities to anticipate others' doings, emotions and intentions. Strong counterfactual simulation seems to be the key ability for such an ability to anticipate. She may on this basis develop a personal conception of lying, being misinformed, and having weird desires. This personal conception will gradually be merged into the common, socially transmitted, folk-psychological theory.

The threat of vicious circularity seems thus to dissolve. For a subject does not need to know herself to know others. She rather needs to evaluate situations in a way both psychologically and rationally similar to other subjects. In the proposed view, as in Gordon's approach, psychological knowledge has nothing to do with an introspective self. The present view stresses further its social nature. Mastering a theory of mind cannot be reached independently of social interaction. But an individual can only benefit from social theory-learning if she is able in practice to represent counterfactual situations dynamically, and to extract from them stable properties.

References

Biro, S. & Russell, J., (2001), The execution of arbitrary procedures by children with autism, *Development and Psychopathology, 13, 97-110.*

Carruthers, P., & Smith, P.K., (eds.), (1996), *Theories of Theories of Mind,* Cambridge, Cambridge University Press.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J. & Jeannerod, M., (1997), Looking for the Agent, an Investigation into Self-consciousness and

Consciousness of the Action in Schizophrenic Patients, *Cognition.* Vol. 65, 71- 86.

Davies, M. & Stone, T., (eds.), (1995), *Folk Psychology: the Theory of Mind Debate,* Oxford, Blackwell.

Davies, M. & Stone, T., (eds.), (1995), *Mental Simulation,* Oxford, Blackwell.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F. & Fazio, F., (1997), Brain activity during observation of action. Influence of action content and subject's strategy, *Brain,* 120, 1763-1777.

Evans, G. (1982), *The Varieties of Reference, Oxford*, Clarendon Press.

Gallese, V. & Goldman, A., (1998), Mirror neurons and the simulation theory of mind reading, *Trends in Neuroscience, 2*, 12, 493-501.

Gergely, G, & Watson, J.S., (1998), Early Social-Emotional Development: Contingency Perception and the Social-Biofeedback Model, in P. Rochat (ed.), *Early Social Cognition,* Hillsdale, N.J.: Erlbaum, 101-136.

Goldman, A. I., (1989), Interpretation psychologized, *Mind and Language, 4,* 161-85.

Goldman, A.I., (1992), In defence of Simulation Theory, *Mind and Language,* 7, 1 & 2 104-119.

Goldman, A.I., (1993), The Psychology of Folk Psychology, *Behavioural and Brain Sciences, 16*, 15-28.

Goldman, A.I., (1995), Empathy, Mind and Morals, in Davies, M. & Stone, T., (eds.), *Mental Simulation,* Oxford, Blackwell, 185-208.

Goldman, A.I., (2001), "Desire, Intention, and the Simulation Theory", in B.F. Malle, L.J. Moses & D.A. Baldwin (eds.), *Intentions and Intentionality, Foundations of Social Cognition*, Cambridge, MIT Press, 207-224.

Gordon, R.M., (1993), Self-ascription of belief and desire, in *Behavioural and Brain Sciences,* 16, 1, 45-6.

Gordon, R.M., (1995), Simulation without introspection or inference from me to you, in Davies, M. & Stone, T., (eds.), Mental *Simulation,* Oxford, Blackwell, 53-67.

Gordon, R.M., (1996), 'Radical' simulationism, in Carruthers, P., & Smith, P.K., (eds.), *Theories of Theories of Mind,* Cambridge, Cambridge University Press, 11-21.

Harris, P., (1992), From simulation to Folk Psychology: The Case for Development, *Mind and Language,* 7, 120-144.

Harris, P., (1993), Pretending and Planning, in S. Baron-Cohen, H. Tager-Flusberg & D. J. Cohen (eds.), *Understanding other Minds: Perspectives from Autism*, Oxford: Oxford University Press.

Harris, P.L., Brown, E., Marriott, C., Whittall, S. & Harmer, S., (1991), Monsters, ghosts and witches: testing the limits of the fantasy-reality distinction in young children, *British Journal of Developmental Psychology,* 9, 105-123.

Heal, J., (1995), How to think about thinking, in Davies, M. & Stone, T., (eds.), Mental *Simulation,* Oxford, Blackwell, 33-52.

Heal, J., (1996), Simulation, theory, and content, in P. Carruthers & P.K. Smith, *Theories of theories of mind,* Cambridge, Cambridge University Press, 75-89.

Hood, B., (1995), Gravity rules for 2-4 year olds. *Cognitive Development,* 10, 375-386.

Jeannerod, M., (1999), The 25th Bartlett lecture; To act or not to act. Perspectives on the representation of actions, *Quarterly Journal of Experimental Psychology, 52A (1), 1-29.*

Leslie, A. M. (1987), Pretence and representation, The origins of 'theory of mind'. *Psychological Review, 94,* 412-426.

Leslie, A. M. & Thaiss, L., (1992), Domain specificity in conceptual development. Neuropsychological evidence from autism, *Cognition,* 43, 225-51.

Papineau, D., (2000), "The evolution of knowledge", in P. Carruthers & A. Chamberlain, (eds.), *Evolution and the human mind, Modularity, language and meta-cognition*, Cambridge, Cambridge University Press, 170-206.

Perner, J., (1991), *Understanding the representational mind,* Cambridge, MIT Press.

Perner, J., (1994), The necessity and impossibility of simulation. *Proceedings of the British Academy*, 29-44.

Perner, J., (1996), Arguments for a simulation-theory mix, In P. Carruthers & P.K. Smith (eds.), *Theories of theories mind,* Cambridge, Cambridge University Press, 90-104.

Povinelli, D., (1996), Chimpanzee theory of mind? The long road to strong inference, in Carruthers, P., & Smith, P.K., (eds.), (1996), *Theories of Theories of Mind,* Cambridge, Cambridge University Press, 293-329.

Proust, J., (2000), "Awareness of Agency: Three Levels of Analysis", in T. Metzinger (ed.), *The Neural Correlates of Consciousness,* Cambridge, MIT Press, 307-324.

Proust, J., (2000), "Can non-human primates read minds?", *Philosophical Topics,* 27, 1, 203-232.

Proust, J., (2002), Imitation et agentivité, in J. Nadel & J. Decety (eds.), *Imitation, Représentations motrices et intentionnalité*, Paris, Presses Universitaires de France.

Quine, W.V.O., (1960), *Word and Object,* Cambridge, MIT Press.

Recanati, F., (1999), *Oratio Obliqua, oratio recta*, Cambridge, MIT Press.

Riggs, K.J., Peterson, D.M., Robinson, E.J., & Mitchell, P., (1998), Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13, 73-90.*

Strawson, P.F., (1959), *Individuals,* London, Methuen and Co.

Trevarthen, C., Kikkinaki, T. & Fiamenghi, G.A., (1999), What infants' imitations communicate: with mothers, with fathers and with peers, in J. Nadel & G. Butterworth, (eds.), *Imitation in infancy,* Cambridge, Cambridge University Press, 127-185.

Tulving, E., (1983), *Elements of Episodic Memory,* New York, Oxford University Press.

Vygostsky, L.S., (1978), *Mind in society*, Cambridge, Harvard University Press.

Woodfield, A., (1996), Which theoretical concepts do children use? *Philosophical Papers,* XXV, 1, 1-20.

Woodfield, A., (1997), Social externalism and conceptual diversity, in J. Preston (ed.), *Thought and Language,* Cambridge, Cambridge University Press, 77-102.

Zaitchik, D. (1990) When representations conflict with reality: The pre-schooler's problem with false belief and 'false' photographs. *Cognition, 35,* 41-68.