Mental Simulation as Affective-Cognitive Act

Otto Paans

> The concept of a representation of what is seen, like that of a copy, is very elastic, and so *together with it* is the concept of what is seen. The two are intimately connected. (Which is not to say they are alike.)
>
> --L. Wittgenstein (2009: 198e)

1. Introduction

Despite a sizable body of theory that has been developed over approximately the last 30 years, the issue of how humans can infer and understand each other's emotions, motivations and reactions remains a perpetual blank spot on the conceptual map. Why are we able to place ourselves in "someone else's shoes" in the first place? The motivations and thoughts of others are not directly readable, save by a recourse to body language and expressions. So, the best information we can acquire about the motivations and behaviour of others stems from observation, or reasoning based on past experience.

One way to explain this human capacity has been to postulate that every human being is equipped with a tacit theory of mind, a position known as "theory-theory."[1] However, postulating a tacit theory or constructing a rule-based system for inferences helps little to clarify the issue, as there is no conceivable limit on the number of logical systems or tacit theories that can be formulated in order to explain the semi-spontaneous grasp that ordinary human beings have on "folk psychology." Whatever theory is postulated, there is no real criterion for choosing between two rival theories.[2]

The other way around the problem is to hypothesize that human beings simulate behaviour of others. This explanation has one advantage over the idea of an innate theory: we can readily imagine ourselves simulating someone else when we say "she must not have liked that remark" or "he will never accept this outcome of the negotiation." In making such remarks about others, we step temporarily into their shoes, and reason from their point of view. We broadly adopt their perspective and think through a given situation from within their (supposed) perspective. Of course,

---

[1] Gordon (1998: 11–21); Botterill (1998: 105); Stich and Nichols (1998: 423); Leslie and German (1995b: 134).

[2] Of course, one could always refer to Ockham's razor on grounds of parsimony, but this would be a criterion that is external in the sense that it is a general preference, and not something that derives from the explanatory power of a theory.

such exercises in mental simulation are as much projections of what we think is important, and are as such laden with value-judgements, blind spots and presuppositions. As heuristic tools, however, such exercises are accurate enough to make us wonder why we have a "semi-spontaneous grasp" of folk-psychological concepts. In this essay, I broadly accept the line of reasoning provided by so-called Simulation Theory (ST) in ascribing our grasp of folk psychology to the capacity of mental simulation.

I will argue the further claim that mental simulation is enabled by the imagination. The peculiar thing about imagination is that we clearly possess it, but that it is quite hard to say what it is. This has led to two contrary conceptions of imagination: either a negative conception, whereby imagining is viewed as idle conjecture, creation of phantasies, or daydreaming. Alternatively, there is a positive conception, whereby imagination is viewed as a capacity that allows human beings to make creative leaps, conduct thought experiments, or entertain conceptual possibilities. This essay will be concerned with this positive conception of imagination.

2. Mental Simulation: A Concise Account and Two Issues

The paradigmatic account of ST has first been proposed by Robert Gordon in "Folk Psychology as Simulation," by Jane Heal in "Replication and Functionalism," and in Alvin Goldman's "Interpretation Psychologized."[3] The core idea advanced in these papers is that predicting another's mental states is accomplished by simulating them, applying a form of "hypothetico-practical reasoning."[4] In all these accounts, the tacit assumption that prediction is the main source of social understanding is at work. The hypotheses that are generated during simulation are means of predicting courses of behaviour, representing the simulator as an agent that is primarily interested in developing new courses of action. In line with this assumption, the idea behind simulation is not just that one imagines being the person who is simulated. Instead, mental simulation proceeds by performing an "egocentric shift": by placing oneself mentally into the position of the person being simulated, one may derive an insight about the motivations of a person against the background of his hypothesized beliefs or background knowledge.[5] If person A simulates person B, the question for A is not "what would *I* do in that situation?" but "what would *B* do in that situation?" To accomplish this shift, one needs to step into B's shoes to such a degree that one's own perspective almost vanishes during the simulation.

---

[3] See: Gordon (1995a); Goldman (1995a); and Heal (1995a).

[4] Gordon (1995a: 64–66).

[5] Gordon (1995b: 55).

One important feature of mental simulation is that it uses a "patched projection" of someone. We never simulate someone else completely, including all his mental and physical dispositions, possible states of mind and reflexes, etc. Gordon explicitly notes that a "total projection" of a person would not yield very reliable results.[6] In its schematic form, a mental simulation uses selected "pretend-states" that are fed into one's processing module, and that generate "pretend-beliefs," that are either fed into one's belief system or decision-system, as depicted below.[7]

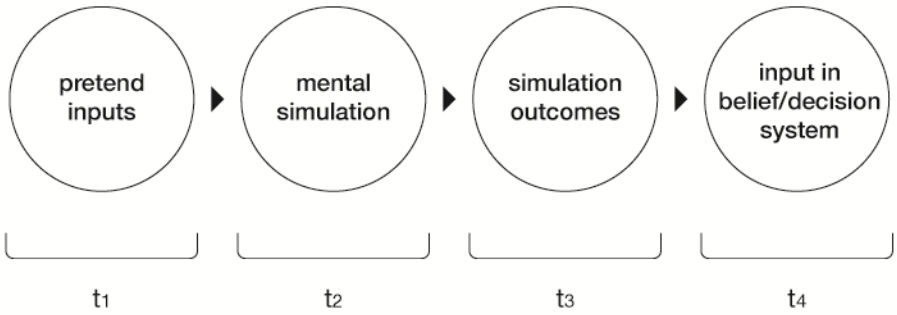

Figure 1: Schematic example of mental simulation: all actions are sequentially ordered, conveying the idea of a split between simulation output and interpretation or input.

This way of representing mental simulation brings to light an important commitment of the early simulation theorists that seems still at work in for instance Goldman's work. The computational (or "boxological") models of simulation seem to favour a process-based approach to cognition that abstracts from bodily experience and affect in favour of clear, sequence-based representations of mental activity. The danger lurking here is that such models are taken as accurate representations of mental activity, allowing one to discuss cognition in isolation of embodiment. Moreover, this way of representing simulation efforts easily overlooks the fact of our embodiment. It represents simulation as a completely mental, brain-based activity. By adopting "boxological" schemes, early simulation theorists like Gordon and Goldman describe simulation in terms already set by their opponents in the theory-theory camp.

Even if we are aware of these limitations and provisionally accept ST as an account of our everyday success in reading and interpreting each other's behaviour, two issues immediately introduce themselves. Let's call them the *Experience Issue* and the *Results Issue*.

---

[6] Gordon (1995a: 102–103).

[7] Stich and Nichols (1995a: 128).

The Experience Issue

What is the relation between former experiences and one's success rate in simulation? In other words, how does background knowledge influence the scope and efficiency of one's simulations? The Experience Issue has received considerable attention in the literature on ST. Gordon has developed an account of ST in which former experiences are not needed at all.[8] Goldman holds that they play a limited role.[9] The thorny issue here is that simulation theorists have not been willing to concede too much: if they would readily accept that the application of background knowledge plays a role in ST, their opponents in the theory-theory camp could all too easily say that they were theory-theorists after all, the so-called "problem of collapse."[10] It would allow TT-proponents to say that simulating is after all a kind of theorizing, and not a different route for representing other minds. If they wanted to push this conclusion further, they could even argue that all mental activity is a subset of theorizing. I do not think this fear well-founded, but it has led ST theorists to develop their ideas on the relation between experience and simulation along defensive lines.

The Results Issue

What can we say about the character of the outcomes of mental simulations? Are those outcomes just new "sentences in one's head"? Can simulation outcomes be described by sentences like: "after running simulation X, subject Y knew that a, b and c were the case for subject Z" – in which a, b and c are new, discrete facts that can be neatly isolated? And, more importantly, what is their epistemic status? Do we know or experience them as facts in the same sense that we know that Berlin is the capital of Germany, or that dark clouds often spell rain? Or are they experienced as guidelines, rules-of-thumb, or simply gut feelings?

Again, the Results Issue has received quite some attention. The underlying problem here can be traced back to the theme of "clear and distinct intuition" as developed by Descartes. The idea is that we can look inwards with the same analysing gaze that we use to look outward.[11] This "Cartesian" introspection is, as it were, a search-light used to illuminate what we know, suggesting a split between possessing a fact and actively realizing that one knows it. If one extends this line of reasoning we arrive at the following view: the mental simulation process is a black box producing results at t_1 that must be separately interpreted at t_2. Therefore, introspection-as-a-search-light is a necessary component for interpretation: the simulator would have to "look inward" at t_2 to see what the simulation produced. In addition, the nature of the

[8] Gordon (2007: 153–154).
[9] Goldman (1995a: 85).
[10] Goldman (1995a: 85).
[11] Heal (1995a: 45); Carruthers (1998: 22); Davies (1996: 112, 122–3); Dennett (1987: 134).

inner states encountered by this searchlight gaze must be immediately clear to him, in the kind of "clear and distinct" insight that Descartes imagined. If we draw a diagram of this process, it is ordered sequentially: its phases succeed each other in time, separating the moment of looking inward and understanding what is being seen. The time between the individual phases may be very short (I might realize "with a shock" that I have been lied to, for example, and this realization may occur in a matter of seconds). Even when we take this into account, the implicit assumption underlying this model is, I think, that a temporal and conceptual split must be introduced between "looking inward" and "understanding what is being seen." The discrete phases depicted in Figure 1 are thus taken literally: the moments of perceiving, understanding, and acting are chronologically and conceptually divided.

The pretend-inputs with which mental simulation start set off the sequence. As response to the pretend states, the simulation starts running, producing outcomes. In turn, these outputs must be interpreted and understood, and are used in inputs in the belief or decision system. Even on first sight, this model seems a little too idealized, and when adhered to too rigidly, produces the erroneous "Cartesian" view.

Some of the attempts to circumvent the Results Issue have been creative, to say the least. For instance, Carruthers argues that the reliability of a statement about one's own mental disposition emerges simultaneously with an associated belief that the perceived state is true. Moreover, this appearance of belief-plus-reliability is in fact strengthened by the fact that it is arrived at non-inferentially.[12] Thus, as far as access is concerned, it does not get more direct than this.

On first sight, this move seems like a very parsimonious theoretical solution. Not only do mental states signal their presence by triggering an associated belief that they are there, but they do so in a way that is actually reliable! The problem of this solution is that it leads to contrary conclusions. Empirical evidence suggests that introspection is in many cases not reliable.[13] The clear and distinct inner gaze of Cartesian introspection seems conspicuously absent. Here, we encounter an explanatory gap: either we can reliably gain access to our own mental states, or our inner gaze is a feeble instrument indeed, in which case Carruthers's theory cannot work, or at least delivers heavily uneven results.[14]

Both the Experience Issue and Results Issue introduce an underlying, even more fundamental question. Even if we could explain the relation between background knowledge and the efficacy of one's simulations, and if we could provide

---

[12] Carruthers (1998: 27); Goldman (1995a: 88–89),

[13] See: Nisbett and Wilson (1977); Dennett (1993); Churchland (1981: 70).

[14] This introduces a new question, namely: how it is possible that some people have a quite well-developed "gut-feeling" or more reliable self-image than others.

an account of the character of simulation outcomes, the problem of arbitrariness would quickly introduce itself. We would have postulated two explanations about different aspects of ST, without explaining how they were related. In other words, the question is whether there is a common ground from which we can explain:

> the relation between background knowledge (or life experience) and the efficacy of the mental simulations that someone runs, thus a response to the Experience Issue, and the character of the outcomes of mental simulations and their epistemic value, thus a response to the Results Issue?

If we could formulate a single phenomenon or capacity that ties (A) and (B) together, the charge of arbitrariness would at least be weakened, as both issues could be solved in a unified account.

I develop such an account of mental simulation in the following section. The argumentation is centred around the capacity of *imagination*. Put concisely, I argue four points:

> (i) that the imagination is a human cognitive-affective capacity that allows for performing mental simulations,
>
> (ii) that because of the functioning of the imagination, we can develop an account of how background knowledge informs mental simulations, avoiding the problem of Cartesian introspection (the *Experience Issue*) altogether,
>
> (iii) that the outcomes of mental simulation produce affective-cognitive states that are not dependent on "sentences in the head," or in need of being interpreted at a later time by a kind of Cartesian introspection or "inner gaze," and
>
> (iv) that this affective-cognitive character has an epistemic value that supersedes the logical coherence and neatness of propositional modes of thought (the *Results Issue*).

I will not argue these points section-by-section, but lay some groundwork before revisiting them in a closing discussion. In the next section, I discuss imagination as an intentional act. This means as an activity that is inherently judgement-laden, containing already conceptual knowledge, affects, or alternatively primitive pattern-recognition that is projected on the world. Following this discussion, I delve into the issue of narrative. Although mental states are most likely not sentences in the head, we obviously have the capacity to *form* sentences in our head. In turn, we can use this capacity to create stories and backgrounds for behaviour we encounter. This capacity

allows us to interpret behaviour in different ways. The imagination allows for creating *coherent* narratives, although it is not just a kind of "sentence-generator" about behaviour. Both features of imagination turn the Cartesian picture upside down, doing away with the need for a clear and distinct inner gaze. Of course, this line of thinking branches out in issues on how infants recognize mental states or the role that concept possession plays in the effectiveness of simulation.

Having provided this foundation, then I will be in a position to explain how mental simulation as enabled by the imagination produces affective-cognitive states. This, in turn, allows me to say more about the issue of access to our own mental states. In turn, this allows me to discuss the epistemic value of such states, thereby tying (A) and (B) together.

3. Imagination as Intentional Act

Right from the introduction of ST onwards, various theorists have appealed to imagination as a fundamental capacity in mental simulation. Gordon, for example, states that one "imaginatively identifies" with the person to be simulated, or that one must make "imaginative adjustments" in simulating other people to account for their specific characteristics.[15] Significantly, the identification with a certain person illuminates the central thrust of Gordon's approach: simulation runs through identification, as the simulator and subject share an analogous cognitive make-up.

Currie has proposed that mental simulation is in fact identical to imagination.[16] Although this might be overstating the case, it points to an ambiguity regarding the function and domain of the imagination: clearly, there is a role for imagination to play in simulation, but what it exactly does needs explication, especially if it replaces introspection-based accounts of third-person and first-person ascriptions.

The term "imagination" refers to a broad range of cognitive and mental phenomena, which makes it hard to discuss it in general terms. "To imagine," means, variously, to daydream, phantasize, reflect, or write a gripping novel. We say of an author that he "imagined" a realistic world, or we imagine that John and Sue might form a couple. In both cases, the word 'imagining' points to the conceptualization of possibilities, but the ways in which this happens are vastly different.

Following Colin McGinn, I adopt a first constraint on the discussion by focusing on mental images. The term "imagining" can refer to forming a mental image (of a beautiful mountain range, say), and it can refer to entertaining conceptual

---

[15] Gordon (1995a: 63); Gordon (1995b: 55–6).
[16] Currie (1995b: 158).

possibilities (imagining an urban expansion on an unused piece of land).[17] In both cases, the output of the imaginative efforts is a *mental image* that can be controlled to some degree. Artists and architects routinely engage in forming mental images and externalizing them in sketches, models, or paintings, in which they exercise a degree of control over what they produce.[18] In a more everyday setting, people imagine how their holiday in Thailand will turn out, or rehearse how a job interview will unfold.

In these cases mentioned here, the mental image is as it were insulated and distinct from the real object. If I imagine a mountain range in Switzerland, my imagination may be influenced by earlier images and impressions, but I do not actually observe the mountain range I imagine. As such, argues McGinn, the image cannot surprise me, as I am its author.[19] To some degree, this claim is true – although not always.

McGinn's argument holds for mental images in the narrow sense, but may stumble to some degree on entertaining conceptual possibilities, especially those cases where the imagination "brackets" usual behaviour, and the notion of pretence enters. Moreover, the imagination may fill in gaps in memory, creating a discrepancy with reality.

For example, when children build a snowman, using an old hat, a carrot, and a stick to finish his appearance, the shaped body of snow is treated in a "pretend mode" as a real man, and its likeness, or his "character" is a matter of pretence. It is treated as a man ("give him a hat, he might be cold"), but in a type of pretend-mode, as none of the children would easily confuse the snowman with a real man.[20]

As the snowman is made, it may be endowed with a background story, special characteristics or a name. The pretend-mode in which the snowman is approached guides pretence behaviour in the same way as belief guides purposive behaviour, although it produces a kind of split simulative experience: the children know that the snowman is not a real man, yet they represent and treat him as such. Their behaviour is simultaneously appropriate to the object as snowman (rubbing on extra snow to make him look more life-like) and the object as real man (giving him a hat or a name). Throughout the game, the clear border between fact and fiction becomes blurred, although some facts may remain consistent with one of the representations. The snowman in its pretended role as real man may receive a hat because he is "cold."[21]

---

[17] McGinn (2009: 595).

[18] Although their visual output may well surprise them, and be better or more successful than they *imagined* – yet another use of the term that falls outside the scope of a mere mental image.

[19] McGinn (2009: 598).

[20] Currie (1995b: 153).

[21] Currie (1995b: 154).

This raises the question how this capacity for pretence and partial attribution of real states is connected to the everyday prediction of behaviour and ascription of mental states. Currie contends that we should seek the answer in the evolved capacity for testing survival strategies:

> Daydreaming and fantasy, along with imaginative involvement with fictions, are made possible by a system that already exists for other purposes: strategy testing. If that is right, the simulator has excess capacity, since consuming fictions has, I suppose, no survival value.[22]

Although this statement sounds plausible from a narrowly naturalistic point of view, it sheds little light on the processes involved in applying imaginative capabilities. Moreover, Currie's account is focused on the notion of pretence, in line with the early theorists of ST, who conceptualized mental simulation processes by postulating pretend-states and pretend-outcomes. In these models, simulation is again thought of as a predictive process that can be discussed in isolation from its embodied manifestation. Simulation seems in such accounts almost reduced to a kind of 'informed ratiocination' about the mental states of others, largely excluding the affective components of the process.[23]

Notably, this account of the simulation process cannot satisfactorily answer the following question that can be posed with regards to the earlier example of mental simulation: if Jean simulates Pierre, what justifies him in taking the simulation outcome seriously as a correct representation of him?[24]

It seems that the observer (Jean) uses a kind of displaced perception to get an insight in Pierre's current mental state. The notion of "displaced perception" says that one may obtain information about phenomenon A by perceiving phenomenon B. For example, when a car driver checks the fuel gauge on the dashboard, he infers from the position of the needle how much fuel he has left. He does not directly observe the fuel tank, but does so via a mediating device.[25] He must assume, however, that the device works correctly, and that it represents the state of affairs in the fuel tank accurately. Moreover, he must be aware of the relation between the position of the needle and the amount of fuel.

---

[22] Currie (1995b: 158).

[23] Nevertheless, this does not seem to apply for Jane Heal's mature theory of mental simulation. Her theory of co-cognition seems to leave room for an embodied account of simulation. See Heal (1998) for an extensive discussion.

[24] Heal (2000: 7–8, n. 7).

[25] Barz (2004: 353).

Applied to mental simulation, Jean must have some connective belief that connects the finding "my simulation of Pierre yields A" to the judgement "therefore, Pierre displays behaviour A for reason D."[26] In other words, some intermediary device, inference, or process must guarantee that the step from simulation outcome to judgement is reliable, or at least plausible. One imaginatively identifies with this more abstract representation of someone else, not with all his features. In this case, the notion of imagination as the production of mental images is already far superseded.[27] While mental imagery in this narrow sense might play a role in imagining the behaviour of others, the simulation results far surpass mere "image production."

To explain why imaginative identification can support connective beliefs between observation and judgement, I will discuss the structure of perception more extensively, and show where we can leave the notion of imagination as production of images behind.

A detailed and slightly different treatment of the application of knowledge on a subject to be simulated can be found in Sartre's work. Sartre states that mental images already imply knowledge:

> If the intention is taken at its origin, which is to say when it springs from our spontaneity, it already implies, no matter how naked and bare it may seem, a certain knowledge: it is, hypothetically, the knowledge (*connaissance*) of Pierre. I admit that this knowledge (*connaissance*) is a simple empty expectation, a direction: in every way it is a direction *towards Pierre*, an expectation *of Pierre*.[28]

Observing Pierre is an intentional act. It is not a neutral glance that happens to touch him, but a glance filled with expectations of him, knowledge about him, and intentions towards him. Jean may see Pierre as his long-time friend, a bit clumsy, socially awkward, or he may know that he is heavily near-sighted or stutters. Thus, the glance with which Jean looks at Pierre is already judgment-laden. Sartre puts it quite boldly when he asserts that 'the image is defined by its intention' – underlining the idea that no image is neutral, that it is always about something, or directed at

---

[26] Barz (2004: 353).

[27] It is worth emphasizing that the term "imagination" I use here has been derived from imagination as image-production. Another possible reading is that the imagination is a creative grasping of the world. Kant, for example, had a rather larger scope for the imagination in mind as he writes: "Imagination is the faculty for representing an object without its presence in intuition,"yet adding the distinction between productive and reproductive imagination. The latter is bound by the laws of association, while the former is "(…) an effect of the understanding on sensibility and its first application (…) to objects of the intuition that is possible for us." (Kant [2009: 256–257]). Accounts of the imagination along Kant's line are – I suspect – easily compatible with the account I sketch here.

[28] Sartre (2004: 57, italics in original).

something or someone.[29] Sartre is not shy about granting the intention an epistemic status. The intention implies a certain knowledge – it is as it were the visible sign that some knowledge about Pierre and his circumstances exists and is applied to the situation in which one happens to find him. The border between "knowing" and "applying" blurs in the formation of the image – a fact that can be easily seen when we think about how hard it is to remove or neutralize preconceptions and stereotypes. Such judgements are so deeply ingrained in the images we routinely form that they form an inextricable component of the image itself.

On Sartre's account, the term "image" refers to quite something else than a mere two-dimensional visual entity. Sartre's images are more like immersive– that is *absorbed, engaged* – mental states, or value-laden projections. Such images are not just viewed, they are *felt* and *lived*. They determine attitudes and intentions towards other human beings.

Imaginative intentions are thereby also saturated with imaginative knowledge since one represents in an image only what one already knows (or thinks to know) to some degree. Conversely, this imaginative knowledge is the result of an act of representing selected features of someone else to oneself.[30]

Therefore, the gaze directed at Pierre is a constitutive glance, not merely a report of Jean's senses that is modified afterwards, but a creative force that actively presents Pierre as a subject with certain characteristics and features.[31] Jean might read Pierre's bumping into someone on the street as a sign of his near-sightedness or clumsiness, depending on the judgements that are included in the glance with which he views him. The background beliefs that Jean entertains about Pierre thus influence the interpretation of his behaviour. In Sartre's terms: Jean *constitutes* Pierre along certain lines, like the children constituted the snowman as a pretended real man with a name and character.[32] Thus, the intentional act of imagination unfolds like a process of narration and commentary, developed along selected themes and features of the person being simulated.

4. Imagination as Narrative Process

Appealing to the efficacy of simulation by appealing to imagination invites immediately one objection: it is always possible for Jean to be wrong about Pierre, and so his false preconceptions about Pierre may become a hindrance to accurately simulating him. If simulation is suffused with imaginative knowledge and prior

[29] Sartre (2004: 57).
[30] Sartre (2004: 69).
[31] Sartre (2004: 69).
[32] Sartre (2004: 69).

judgements, it is cognitively penetrable. Thus, if Jean's simulation runs not in isolation from any prior knowledge, the imaginative identification in simulation may be only a way of smuggling in existing biases, an issue known as cognitive penetrability.[33]

To this objection, Sartre replies that it rests on a mistaken view of the mental image: The image is not a representation of reality in the same way that a photograph statically depicts a state of affairs in the past. Instead, during simulation, the faculty of imagination is engaged in an active process that keeps track of changes in the person being simulated.[34] Knowledge about others represents on this account "the active structure of the imaging consciousness."[35]

This so-called active structure is the seamless connection of different parts of a narrative in a whole. To use Sartre's example, when one reads in a novel that Pierre has an argument with a friend in his office in the suburbs of Paris, we imagine the whole office, raised voices, slamming with doors etc. We do not just imagine the components of the story in a loose and unconnected manner. When reading the novel, its storyline does not appear to us as a string of loose, unconnected terms like "Pierre," "suburbs," "argument," "office," "Paris," etc., but as a meaningful connection of those terms.[36]

This move by Sartre at least provisionally provides an answer why one would turn to narrative to understand our social interactions. The mental images we create are suffused with judgement. Whether those narratives and judgements are accurate representations of other minds is a different matter. What is important in this account is that the images themselves are continuously created, setting off a stream of feedback that is used to understand behaviour of others and to modify one's own. Admittedly, a restriction applies here for infants and possibly animals with well-developed mental capacities, such as apes or dolphins. It seems plausible (see below) that at least part of the narrative or response formation takes places unconsciously or in the absence of well-defined concepts that are present in adult human persons.[37]

---

[33] See: Stich and Nichols (19970; Saxe (2005: 178).

[34] Sartre (2004: 61–62).

[35] Sartre (2004: 61).

[36] Sartre (2004: 64–65).

[37] Although the space is lacking to work out a philosophical account of concept possession and formation here, I would like to remark that a sharp distinction between conceptualism/non-conceptualism can be questioned. One reason touching on the discussion can be summarily outlined here. I would say that concepts are not the type of entities one possesses in the same way that one possesses an object or can recall a fact. Moreover, it seems at least plausible to me that infants possess a set of innate skills for pattern-recognition that allows for social cognition and that crystallize into concepts later in life. In-between non-conceptualism and conceptualism there would be a kind of twilight zone of "proto-conceptualism" in which a subject interacts by applying basic (proto) concepts that are in their infant stage, as it were.

The objector may insist that, even if we admit that we effortlessly create storylines and narratives, this response accounts only for how individual sense impressions are being connected and processed. The reliability of those "edited" judgements and resulting beliefs is still a matter of dispute. How does introducing the notion of imagination as the core capacity that underlies mental simulation dispel the issues regarding the reliability of simulation? Put differently, how can questions about the justification of ascribed mental states to others be answered? Even if one admits that simulating is not a perfectly accurate mechanism for prediction by allowing for shortcomings in everyday performance, the doubts may remain.

To respond to this question, two further, connected aspects of simulation and the theory-laden character of perception must be introduced. First, the human capacity to single out individual characteristics of objects. Second, the ability to construct background stories for their existence or the way they co-exist in an object or person. Wittgenstein dedicated a sequence of remarks to these interpretive capacities, noting that when one sees a triangle, a wide variety of judgements can be made about it.[38] It can be 'read' as pointing upwards, as being fallen over, as hanging from its apex, as being sharp etc. In each of these cases, a background narrative about the triangle is constructed. The content of the narrative goes beyond the observable characteristics of the elements of the image. After all, the image consists only of three straight lines jointly forming a set of three corners whose sum is always 180 degrees. Yet, the narrative provides a reading or interpretation of these elements, and furnishes them with a meaning. In each of those readings, one aspect is singled out for playing a key role in the interpretation: the triangle is interpreted as *resting* on its longest side, or its apex as read as *pointing* in a certain direction. Furthermore, it is possible to switch between narratives, seeing the triangle alternately as pointing up or hanging from its apex, just like the perceptual shift in the famous duck-rabbit illusion. In each of the cases, a certain aspect "lights up."[39]

Wittgenstein's observations echoe Currie's example of children playing with a snowman: the snowman is imagined as a man (or, to use Sartre's terminology, constituted as a man), and receives its shape because of the background story. The background narrative directs behavioural dispositions towards the roughly shaped heap of snow with two buttons, a hat, and a carrot. Juxtaposed without a connecting story, these elements are meaningless. To overcome this meaninglessness, a kind of integrative, narrative act provides all elements with a common meaning that can be interpreted, and that consequently guides further behaviour.[40] Wittgenstein contends

---

[38] Wittgenstein (2009: 210–212, and 219).

[39] Wittgenstein (2009: 204).

[40] The interesting ambiguity here is that Wittgenstein in (1980: vol. 2, §102 – §105) opens up a related issue, namely whether I have the same relationship to my own words as to those spoken by others.

that reading objects (and persons) in this way is determined by an attitude towards them. "It is," he writes, "as if an idea came into contact with a visual impression."[41]

In mental simulation, we use a variation on this process (or a very similar process), as may be illustrated with the case of an actor. In preparation for his performance, an actor imagines his stage persona: his character, his background, the way he walks, whether he is brusque, gentle, or indecisive... This preparation allows the actor to take the acting performance beyond the script – adding depth and life to the lines and phrases, just like the background narrative provides a meaning to the individual elements of the snowman.

Some actors may be completely unconvincing as a villain, as loving husband, or as lawyer. From the point of view of the audience, there must be some congruence between the perception of the character and the dramatic characterization presented by the actor. This is Sartre's point: when someone acts, he plays a role, although we still recognize the natural person underneath the role if we want to.[42] A strange duality ensues: perhaps the actor convinces us through his performance that he really is a villain capable of evil, yet we also know he is a decent person in real life. Just as in the case of the snowman, the capacity for pretence allows human beings to entertain a double, simultaneously existing image of other persons. On one side, there is the other person as observed through his physical appearance and behaviour; and on the other side is the image through which he is constituted by an observer. Barz notes that "taking part in games of make-believe means having a split consciousness": while focusing on the fictional facts, the participant is aware of the real facts at the same time. "[One] never gets a 'pure' glimpse of the real objects, since the props block [one's] view."[43] The character played by the actor is a prop that disguises the real person for the time being, allowing the observer to view the actor as someone else.

According to Barz, we use the actor as "stand-in" for the character. This allows us to pretend, for the duration of the theatrical performance, that the actors are really their characters. In similar fashion, Jean's projection of Pierre is a stand-in: in effect, Jean believes for the duration of the simulation that Pierre is to some degree equivalent to his "patched projection" of him. This skill of using stand-ins allows us to create fictional worlds with associated fictional truths. In such worlds, we can pretend that "X is true" (relative to the rules of that world) while simultaneously knowing it is not

---

This is of course a related issue, as the same narrative spoken by me or by someone else may solicit very different affects and may cause me to draw very different inferences from them.

[41] Wittgenstein 2009: 215–7

[42] Sartre 2004: 28. Although not always. An interesting case is the career of Dutch actor Erik de Vogel, who plays a villain (Ludo Sanders) in a long running Dutch soap opera series, and has been confronted multiple times in real-life by viewers who really thought that he cheated on his wife last night…

[43] Barz (2014: 360).

true in the real world.[44] To maintain this split consciousness in simulation demands imagination to superimpose the fictional world on the real world, while deriving information from the former to apply in the latter.

If we take this thought one step further, we can propose that simulation uses internal props, by taking our visual or imagined experiences temporarily as "fictionally true," and thereby leading to a kind of "what if" reasoning, opening up the way for the imagination-driven scenario thinking.[45] One might simulate someone else even in his absence, indulging oneself in all kinds of predictions about how he would react in certain circumstances.

Likewise, the character played by the actor may invoke feelings of sympathy, disgust, or anger, in addition to speculation what his character would do next. In this way, literature, theatre, and movies are powerfully immersive, because they stimulate and direct the process of sympathizing with or simulating the characters. As Sartre puts it:

> To read a novel is to take a general attitude of consciousness: this attitude largely resembles that of a spectator, who, in the theatre, sees the curtain rising. It is preparing to discover a whole world, which is not that of perception, but neither is it that of the mental images. To be present at a play is to apprehend the characters *on* the actors, the forest of *As You Like It on* the cardboard trees. To read is to realize contact with the irreal world *on* the signs.[46]

Just as in the case of the triangle, the observable and describable physical characteristics of the actors and the stage are one half of the coin, the other half busily filled in by the spectators on the basis of clues in the text, background information in the story, the acting performance, or the attitude they adopt towards the characters. Wittgenstein observes the same duality when he notes that when persons describe what they perceive, they do not need to think separately about the object they are describing. The visual impression has as it were a direct connection with the thought expression.[47] The imaginative component (seeing the cardboard trees *as* a real forest) is included in the thought expression, "This scene takes place in the forest."

Thus, imagination is simultaneously an intentional act *and* a narrative process. When we perceive others by means of imaginative simulation in the world around us, we do so in a way that is irreducibly laden with judgements, expectations and knowledge. This combination of intentionality and narrating capacity enables people

---

[44] Barz (2014: 358).
[45] Barz (2014: 359).
[46] Sartre (2004: 64).
[47] Wittgenstein (2009: 207, §139).

to create background stories on the events and behaviour they perceive in their surroundings. However, simulating others is not just a matter of ratiocinating about them. The actor on the stage *affects* the spectators. By displaying certain behaviour, he triggers emotional states in his audience that are felt as well as rationally known.

5. Imagination as Affective-Cognitive Capacity

The same combination of affect and cognition (although not so clear cut as in the theatre setting) may be applied to mental simulation. In the example of Jean and Pierre, it was possible for Jean to simulate Pierre through patched projection: certain characteristics of Pierre were incorporated into Jean's simulation, while others were downplayed or altogether left out. This "new Pierre" is an incomplete subject, as he consists of selected factors and characteristics that Jean (according to his imaginative skill or the situation in which they both were) thought relevant to ascribe to him. In simulating Pierre, Jean adopts an attitude towards him, allowing him to highlight certain aspects and providing a different reading of Pierre's behavior. Just as one "takes a general attitude" when reading a novel or watching a movie, so too does simulation rest on attitudinal dispositions.[48] In the theatre, this process works precisely the other way around: an actor exaggerates certain aspects of his character to provide the audience with a certain image of him, a kind of readymade patched projection.

The outcomes or results from mental simulation are not only informative in the sense of merely providing Jean with more knowledge, or the audience with a clear image of the actor. They are also simultaneously affective. This means that they are felt as well as known.[49] If Jean imagines how he would feel if he were to be followed by a stranger through a dark alley, he may not only know that he would be afraid, but could also simultaneously feel it. His breath would become shallower, or his heartbeat might rise. The affective component occurs in parallel with the realization, "I would be scared in that situation." Unlike the Cartesian accounts of introspection, John is not "waiting" for a result to appear after the simulation has taken place.[50] The result is immediately present in an affective form, coupled with a realization that can or cannot be expressed in a proposition. (Recall a situation in which someone tells you, "I did not know why, but it did not *feel* right").

In the case of the actor, his performance seduces the audience to "feel" or "tune in" on his behalf. If he is a villain, the audience may start to dislike him. If he is a victim of injustice, the audience may empathize or sympathize with his situation.

---

[48] In this connection, a question can be raised whether they emerge voluntarily, involuntarily, neither or both. However, this question is outside what I wish to discuss here.

[49] Currie (1995b: 158); Sartre (2004: 28, 68–69); Gordon (1986); Machery (2005: 473).

[50] Admittedly, not all acts of the imagination are that clearly affect-saturated.

Sartre argues on this point that persons or characters are perceived first as "representative structures," that is, objects or entities that can be described by listing their empirically observable properties.[51] In the case of mental simulation, we may perceive someone as tall, short, blond, dark-haired, etc. In the second instance, however, these persons or characters are read in such a way that on top of the describable characteristics, they are presented within an affective structure.[52] This means that encountering someone else is not just a matter of listing all his observable characteristics. Instead, these features jointly constitute more than the sum of their parts, namely a person or character for which sympathy or disgust can be felt, and at whom anger or indignation can be directed. Sartre's account of superimposing representative and affective structures parallels a central idea in the philosophy of Hume, namely the concept of *sympathy:*

> A good natur'd man finds himself in an instant of the same humour with his company; and even the proudest and most surly take a tincture from their countrymen and acquaintance. A cheerful countenance infuses a sensible complacency and serenity into my mind; as an angry or sorrowful one throws a sudden damp upon me.[53]

In this passage, Hume underlines something fundamental to human interaction: namely, the fact that mental states are not only known by ratiocination, but directly and simultaneously by felt experience. Hume's account of sympathy stresses the fact that the states of others are not only known by observing behaviour or theorizing on patterns or causes, but by an affective form of communication that exceeds pure ratiocination or inference:

> Hatred, resentment, esteem, love, courage, mirth and melancholy; all these passions I feel more from communication than from my own natural temper and disposition.[54]

The affective structure gives other persons a personality or character, allowing them to influence others. Moreover, affectivity causes human beings to view others as a proper object for simulation, the assumption being that the other person is in relevant aspects like oneself, or whose motivations and reasons can at least be understood to some degree.[55] (The question that introduces itself here is whether we not unwittingly project our own reasons and preferences on our interpretation of

---

[51] Sartre (2004: 70).
[52] Sartre (2004: 70).
[53] Hume (1739: 318).
[54] Hume (1739: 318).
[55] Gordon (1995a: 114–120); Hume (1739: 319); Dennett calls this "the intentional stance."

others. If this is so, then even our affects are cognitively penetrable, and are liable to be influenced by our mood, emotions, preferences, etc.).

The affective capacity of performance is fully exploited in acting and writing literature. The characters do things we can imagine happening, and to which we can respond affectively. We do not merely ratiocinate about such characters, but their performance includes an element of sharing: we feel what the characters in a play feel. We may cringe at betrayal, or sigh with relief if two lovers in a movie are united again.

To see what the relation between imagination and affective communication between humans exactly entails, a short explanation of Hume's theory of ideas and impressions is necessary.

Hume maintains that imagination is the human capacity that makes the transition between idea and impression, giving the impression its affective force.[56] (The idea, according to Hume, is not different in content from an impression, but has a lesser force and vivacity, or less affective impact).[57] Imagination is the bridge between the rational and affective components of an impression. Without imagination, we would end up with the formal world of pure ratiocination about behaviour (perhaps, to some degree, this is what people with Asperger's Syndrome experience). In such a world, it would be hard to see how we could develop empathy, friendship, or pity. Hume explains how one can be overcome with a feeling of empathy, kindness, or friendship when looking at the portrait of a friend.[58] The portrait serves as a prop that one may use to activate the affective feelings by invoking a visual image referring to someone towards whom one entertains feelings of kindness and friendship.

Therefore, the imagination plays a key role in understanding other persons through its affective force. When sympathy is added to human interaction, the motivations and reasons of others are first inferred by reading behaviour and "external signs." Then the ideas stemming from this reading are converted in impressions, and "acquire such force and vivacity" that they produce the same affective effect in the observer.[59]

Importantly, as the examples with the portrait or the convincing performance of an actor show, ideas can be transformed into impressions by using props like portraits or performances. The idea of a ruthless villain is infused with an affective, emotional impact once the villain is played by a convincing actor who brings the

---

[56] Hume (1739: 319).
[57] Garrett (2005).
[58] Hume (1777: 52).
[59] Hume (1739: 318).

character to life. In a similar fashion, mental simulation is the capacity enabling one to step into the shoes of someone else, not merely by considering his reasons or desires along rational lines, but by experiencing some of the emotions or feelings that motivate his reasons or desires.

Cognitive science has proved Hume right on this point: brain imaging studies show that when someone observes another person performing an action A, the observer activates a so-called dynamic representation of the corresponding action in the neural structure normally used to execute the action.[60] This means that active "coding" of the action in the brain takes place both by executing and by observing it. Again, Hume has been remarkably prescient:

> Affections readily pass from one person to another, and beget correspondent movements in every human creature.[61]

This communicative process is based on neuronal mirroring activity in the brain that forms a bridge between observing, mentally enacting, affectively experiencing, and remembering. According to Joelle Proust, simulation is enabled by a type of dynamic memory process used to codify procedural knowledge, but cannot be divorced from conceptual capacities. The combination of neuronally mirroring observed actions and simultaneously understanding them on a conceptual level helps subjects to categorize and infer responses from external events.[62] Thus, simulation is dependent on affective-cognitive states marked by their experienced quality, about which one may rationally think and reason. The fact that such states are affective does not place them in a domain beyond reason. However, some affects cannot be directly put in intelligible terms. The person who left the building because "something did not feel right" might lack a discursive explanation about where this feeling came from, yet it prompts him to action, and he can recall it as a motivation for doing something.

In everyday human interactions, the judgement-laden character of perception ensures that others are not only perceived as a series of characteristics that can be empirically described, but at the same time as an object of affection. In other words, we view other human beings as subjects, not merely as objects. Sartre notes that if one would remove the affect, the world would be "singularly impoverished."[63] It would reduce human interaction to pure ratiocination without any emotional or affective component informing it.

---

[60] Proust (2002: 209).
[61] Hume (1739: 577).
[62] Proust (2002: 210).
[63] Sartre (2004: 69).

Summarizing the line of reasoning sketched in this section, the imagination is a capacity that allows for having *affective-cognitive* mental states that are simultaneously comprised of rationally known facts and experienced, emotional states. As Wittgenstein notes, what is derived from an observation, or which aspects "light up" depends on one's attitude to the object of observation. Here again is a clear link with Goldman's (and to a lesser degree Gordon's) account of ST: simulation is a projective activity, in which background assumptions and information are projected on new observations.[64] Conversely, this projection happens almost seamlessly: In the example of the theatre, the audience has to do no conscious effort to see the actors as characters. This way of perceiving the actors is as it were included in the notion of "going to the theatre."

At this point, the skeptical objector might still not be persuaded. He could say, for example, that even if I empathize with a character in a novel, a colleague of mine, or a couple in a movie, the affects I feel are merely my emotions about what I encounter. As such, these emotions may be replete with preferences, personal judgments, and biases. Therefore, they are reflections of my take on the situation, not an accurate representation of reality itself. In the case of reading a novel or watching a movie, this is not important, but when trying to estimate the responses of someone else, this shortcoming could have serious consequences.

This objection gives rise to two separate questions. First, how can I be certain that the affects I experience are indeed about the situation I encountered, and are not just a product of my take on the situation? Second, even if the imagination is a capacity that accounts for the fact that we can empathize with others, how can we know that our ascriptions or predictions are to some degree reliable? I will address the first question in this section, and the second question in the next section.

In response to the first question, we can state that someone's subjective take on a situation is not problematic in a strong sense. Even if one errs now and then in predicting or estimating the motives or behaviour of others, this does not prove that therefore all experiences gained through simulation are untrustworthy. Moreover, the question that the skeptic asks here is about being fully objective towards feelings that are subjectively experienced. However, this betrays a presupposition on the part of the skeptic: to obtain information via a kind of disembodied, direct gaze that is not mediated by our embodiment. Our bodies mediate our experience (such as emotions and affects), and this seems to me an irreducible fact. The directness that the skeptic demands is not possible without annulling our embodied existence.

---

[64] Gordon (1995a: 107; Goldman 1995a: 84).

6. Imagination as Enabler for Mental Simulation

Still, there is room for a further objection. In other words, even if we appeal to the imagination as the capacity that allows for having affective-cognitive states and as fundamental feature of mental simulation, is there any plausible explanation for regarding the outcomes it produces as reliable, or to view them as justifiable? Is appealing to imagination capable of providing a degree of reliability that an account focused on Cartesian introspectionism cannot?

To provide a response to this objection, we have to return to Hume. As discussed in the previous section, Hume has been proven partially right by cognitive science when he noticed that we can infer the states of others by a form of affective communication. Observing certain actions being performed caused the brains of observers to re-enact the action observed. This effect is called "mental mirroring," and its existence may be used to argue for the plausibility of mental simulation.[65] Hume himself could not foresee how prescient his remarks would turn out to be, because he discussed human interaction, not the neuronal structure of the brain.

Goldman has further developed ST based on these more recent findings in neuroscience, and follows a naturalistic line of reasoning. He distinguishes between two stages of mental simulation: the first one is mirroring (based on mental mirroring), the second one is mindreading.

Mirroring is a neuronal activity of the brain that occurs when someone observes certain actions being performed. Its occurrence has been used as an argument to defend mental simulation. In a study on this phenomenon, test subjects were shown pictures of hands and feet that were contorted in seemingly painful positions. They were asked to specify the intensity of the pain they could "read in" the images. In all cases, both the affective and sensorimotor parts of the pain system were affected while the subjects tried to answer the question. The sensorimotor part would for example cause someone to wince "in pain" on seeing someone else's body parts in hurtful positions.[66]

Mindreading is the stage of actual attribution, prediction, or explanation of mental states based on inference and felt experience. Mirroring happens pre-reflectively, as it is a process that is automatically triggered in the brain.[67] Conversely,

---

[65] See, e.g., Gallese and Goldman (1998).

[66] Goldman (2009: 245).

[67] There are a number of points in Goldman's overall account I disagree with. The most obvious one is his attempt to reduce embodied experience to neuronal mechanisms, hence reducing lived experiences to blind mechanisms, in keeping with materialist/physicalist explanations. However, the

mindreading is a purposive attempt to explain or predict behaviour. However, mindreading may to some degree be carried out without explicitly noticing it:

> It is then argued that there are two forms of mental simulation. The first is mirroring, which is automatic, almost entirely unconscious, typically involves comparatively "primitive" mental states, and doesn't rely on task-specific knowledge or information. This kind of simulation, called "low-level simulation," is to be distinguished from "high-level simulation," which corresponds to the original idea of pretence-driven, or imagination-driven, simulation. High-level simulation is more effortful, is sometimes conscious, characteristically involves more complex mental states, and is guided by task-specific knowledge or information.[68]

Goldman identifies mirroring as the causal explanation for mindreading. The mirror stage has to take place to be able to "read" someone else's intentions at all. Thus, the claim is that mirroring is a preparatory stage for actual mindreading: mirroring (what Hume would possibly have called "sympathy") is a necessary condition for the attribution of mental states. Before we attribute a mental state, we must "feel" it from the inside, allowing us to sympathize with the other person. This affective feeling provides the attribution not with an infallible theoretical status, but the fact that it has an affective component gives it a very special type of authority. The attribution is based on extrapolation from what is felt in one's own case, projected on someone else through mental simulation. In other words, the affect-suffused perceptions one experiences are an epistemically distinguished class of experiences. So this account of attribution is not just merely focused on ratiocination following the rules of logic, but even more fundamentally on affectivity.

The epistemic authority of simulation derives from the fact that affective feelings evoked in mental simulation play an informative role alongside facts and ratiocination. This combination provides insights or tentative explanations about what motivates the behaviour of others. The fact that we can derive information from others by ratiocinating about them does not entail that only facts or deductions about the behaviour of others are informative. Instead, the affectivity of mental simulation may even enable swift selection of plausible explanations or predictions of behaviour. The reliability of simulation is thus strengthened by the fact that the imagination plays a central role.

Whether the two levels of simulation can be sharply distinguished can be debated, but the distinction that Goldman draws, namely between an unconscious

---

distinction he makes here seems useful to me, even if I do not agree with the consequences which Goldman derives from it.

[68] Goldman (2009: 246).

(and non-conceptual) resonance ("sympathy") between persons and conscious, imagination-driven scenario thinking seems feasible, as it broadly corresponds to the affective and cognitive aspects of the imagination.

Roughly, Goldman's approach parallels Hume's argument that ideas are converted into impressions. First, observations are used to infer causes of behaviour, and in a second step, these causes are internalized and affectively experienced. The capacity of the imagination (for example imagining someone is in terrible pain) causes the transition from mere observation or fact to affectively experienced mental state. In turn, these mental states enable a second type of simulation, namely mindreading.

Hume's account of the imagination recounts how impressions can be freely (although associatively) combined by the imagination, but that this process is done in a purposive, yet fleeting and creative way.[69] Impressions that are being experienced (for example, seeing hands and feet in painful, cramped positions) can be freely used by the imagination to work out whether the perceived situation matches earlier impressions stored in memory.

Therefore, the process that Goldman calls "mindreading" has a high degree of accuracy because it combines empathy with rational scenario-thinking capacities. The upshot of this idea is that simulation combines the felt affect of a situation with reasoning capacities. It is a form of empathic scenario-thinking that derives its usefulness not just from theoretical acumen or skilfully applying the rules of logic, but also from the fact that it derives its starting assumptions from mirroring others.

The "force and vivacity" or affectivity of this mirroring process allows for framing the observed behaviour from others in ways that are relevant to the situation. On this account, mindreading also provides on this account an answer to the frame problem: the (pre-reflective) executed mirroring process is a first heuristic device for selecting a bandwidth of scenarios that appear plausible for the target being simulated.[70] This selection process is followed by a phase of mindreading in which these approximations are further refined.

6. An Imagination-Driven Account of Mental Simulation

Tying the various distinct strands of my argument together now, I can formulate a single multi-part argument for mental simulation as driven by imagination.

---

[69] Hume (1739: 9–11) and (1740: 662).

[70] A further question opened up here is whether the mirroring process is non-conceptual, or not. It may be a blind, neuronal process, and in that case the distinction conceptual/non-conceptual may be not applicable at all.

The imagination enables thinking in scenarios by utilizing patched projection to select the most relevant properties for a given situation. Patched projection can be viewed as an economic response to the frame problem, especially with regard to its representational aspects: by selecting several key properties, the subject to be simulated is actively constituted in a certain way, effectively representing a limited body of information about him within the context of a mental simulation. Thus, the fact that Jean entertains a deliberately modified picture of Pierre enables him to simulate him quickly, generating predictive or explanatory scenarios of the type "when Pierre does A, this might mean B, C, or D", "knowing Pierre, he would not have done X without reason Y." Conversely, the imagination allows us to take our visual experiences temporarily as fictional truths in a game of make-believe. Jean uses Pierre as a "prop" for projecting his patched projection, interpreting Pierre's actions through the lens of the perspective he adopted. This line of thought closely resembles Sartre's account of imagination as an intentional, directed act.

The imagination allows one to constitute others as subjects with certain characteristics, accepting this view of others temporarily as true or at least representative. This constitutive glance allows one to create narratives around the behaviour and motivations of others that can be woven into largely coherent (if sometimes mistaken) wholes. The human capacity to single out individual aspects in favour of others, allows one to create different narratives that can be used as helpful tools in explaining or predicting behaviour.

This raises the question how one can effectively select relevant information to create a patched projection. The answer is here partially provided by Hume: judgements regarding the mental states or behaviour of others are not just arrived at through ratiocination, but also by the fact that they are experienced affectively, through what Hume calls "sympathy." If one experiences affectively what someone else experiences by simulating him, a mental impression is created.

This impression has a distinct affective-cognitive character. For example, not only do I rationally know that someone is in pain, but I feel simultaneously what it must be like to be in pain. The mental mirroring process activates both affective and sensorimotor parts of the pain system. The mirroring activity of the brain may cause people to remark, "Ouch, that must have hurt," and making a face when they see someone stub their toe. The impression allows one to proceed from mirroring the observed situation to engaging in directed mindreading about the degree of pain, but consequently also to explaining or predicting behaviour.

In other words: the impression allows one to engage in attribution of reasons and the making of predictions with a high degree of fidelity, the reason why Goldman

calls it "mindreading" – underlining its uncanny accuracy. The fidelity of this process stems from the fact that it is both affectively experienced *and* rationally known.

The imagination facilitates the transformation from impression to idea, if we follow Hume's line of thought. Put in contemporary terms, imagination allows for applying structural features of earlier experiences (the original impressions) to current situations by using abstract ideas (for example, "what it is like to be in pain"). This usage of abstract ideas enables us to postulate an account of mental simulation that allows for effective application of existing background beliefs in simulating.

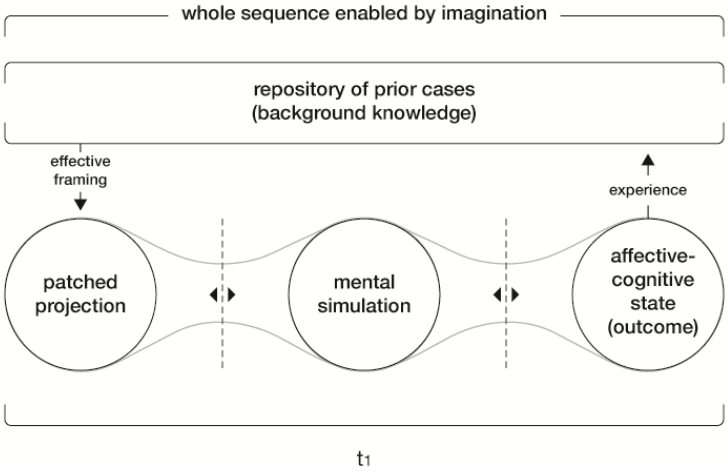

Figure 2: Simulation process driven by imagination-driven mental simulation.

Hume's theory supports on this point Goldman's account of ST, whereby earlier experiences and simulation outcomes are included and effectively applied in mental simulation. The body of information used as theoretical background is not organized like a proto-scientific theory, but as a loosely organized and easily retrievable repository of case studies. Far from being a disadvantage, the fact that simulation is cognitively penetrated by earlier impressions and ideas turns out to be advantageous: by applying affective and rational knowledge from the past through simulating, one can gradually become better at predicting and explaining behaviour of others. The fact that the imagination enables people to have affective-cognitive states adds to the fidelity of the simulation process.

When we take this thought one step further, we cannot just conclude that with regards to fidelity, there are no grounds to make simulation dependent on direct introspection, but even that simulation is better positioned once it is decoupled from introspection. The imagination seems to be a better overall candidate to serve as grounding of our simulative capacities, as it avoids problems of reliability and direct access that inevitably follow once we cling to the idea of Cartesian introspection.

Revisiting the two issues introduced earlier, we can conclude that imagination-driven simulation does away with the need for direct, Cartesian introspection to interpret the results of mental simulation. The "images" we create of others are not results from a separate process, but are interwoven with both ratiocination and empathy. Wittgenstein's ideas about the interconnectedness of what is represented and what is seen are therefore quite near the mark. Both are extremely elastic, and can be influenced by the way we look at the world. Consequently, images (or broadly visual representations) are the result of an attitude towards the world, not merely of a detached, objective gaze. The *Results Issue* can thus be avoided by rejecting the distinction that gave rise to it, pointing to the functioning of the imagination.

The imagination allows human beings to actively constitute others along narrative lines that they choose and develop, based on a repository of experiences from earlier cases. This process has a high – although not infallible – degree of fidelity, since it is affective-cognitive in character. Imagination-driven simulation arrives at its judgements by effectively combining rational conjecture and affective experience, allowing for accurate prediction of human behaviour. The *Experience Issue* can at least provisionally be answered by postulating affective-cognitive states that are known as well as felt.

This approach is well in line with content non-conceptualism, namely, the idea that not only is the content of cognition not always determined by our conceptual capacities, but also perceptual states can have a content that is fundamentally different from the content of a propositional attitude.[71] According to content non-conceptualism, then, sense perceiving can happen without mediation by clear, well-determined concepts. Moreover, perceptual states can have an affective character. One can "feel the pain" when imagining or witnessing someone stumble and take a fall. Hume identified this capacity with the generic notion of "sympathy", but is unwittingly very close to the non-conceptualist split between the content of perceptual states and propositional attitudes. He chooses to draw the line slightly differently, namely between ideas and impressions. Yet, the same insight is at work, namely that human cognition need not be reduced to concepts.

To conclude, this approach does away with a few tenets of contemporary thinking about mental states. The idea that mental states like "belief," "hope," "desire," "excitement," etc., are monolithic entities that belong to the same conceptual domain (or at least can be spoken of as if they did) is rejected by this approach. The fact that we are embodied minds and living organisms obliterates the strict border between "knowing," "doing," and "feeling." This is not to say that there are no unambiguous cases in which people can be said to know or to feel something, but it

[71] For a definition of the term "content non-conceptualism," see Bermúdez and Cahen (2015).

recognizes that exactly these cases are idealizations that cannot be indiscriminately projected on all mental activity.

Much more can be said on the topic of the experiences themselves, of course. For instance, would it be possible or even desirable to phrase such experiences in propositional terms? Can we still think in terms of "truth" or "falsity" when thinking about affective-cognitive mental states? Or should we resort more to terms of "rightness" or "fit"? What kind of mental dispositions are "knowing" or "realizing" on this account?

At this point, one two-part fundamental issue needs to be worked out in some detail: where does ST stand in relation to the wider field of social cognition, and how does the account I have worked out here improve on older versions of ST?

7. Imagination-Driven Simulation and Social Cognition: An Encounter

First, let's shortly recapitulate the main aims and points of this essay. Its main aim is to argue that imagination is a basic human capacity underlying our proficiency in representing other minds by means of mental simulation. The central point is that imagination is an irreducibly cognitive-affective capacity that does away with the need for "Cartesian introspection." Since the imagination is a cognitive-affective capacity, its results are directly, bodily and emotionally felt as well as rationally known. This gives mental simulation its epistemic authority, and is why it cannot be decoupled from embodiment.

On this point an additional question introduces itself: is mental simulation as developed by Heal, Gordon, Goldman, and Currie, also our primary mode of understanding one another? In other words (borrowing a formulation from Gallagher) is mental simulation equivalent to social cognition?

If mental simulation is indeed the primary capacity that allows us to understand one another, how then can infants display a degree of understanding? Gallagher notes that infants can recognize faces, followed in a later stage by the understanding of gestures or expressions. Later still, infants can situate others in a broader world. The picture of the world as we know it as adults comes slowly into being in the infant's experiences.

A second question introduces itself here. If narrative and an observational stance play such important roles in simulation, how do infants navigate the social world around them? In the absence of narratives, how can they understand or grasp mental states of others? The problem, as Gallagher puts it, is that

> [t]he kind of inferential or simulation processes found in explicit versions of TT and ST are too cognitively complex to account for the infant's ability to understand the intentions of others.[72]

This seems to be undeniable. However, further on in the same article, I think Gallagher unwittingly shows the two issues that are conflated here. He quotes Wittgenstein as saying that we simply *see* things in the face of someone else.[73] We do not engage in all kinds of narratives or scenario-thinking to see that someone else is for example afraid or disappointed. This seems also absolutely true. It is possible to infer a vast range of emotions from just seeing someone's face or observing one's body language. Infants also have this capacity, and through experience, they get better and better at it.

Nevertheless, there is a distinction to be made between perceiving what someone else's emotion is (recall the example of seeing someone stub his toe) and understanding someone's motives or desires. The distinction to be made is between what is *observed* and what is *understood*.

In some everyday cases, the perception of what is the case and the understanding of why it is so (or must be so) completely overlaps. One does not need superhuman computational or predictive powers to see one's colleague stubbing his toe, making a face, and saying "Ouch!," and then understanding that this colleague must be in pain. One could easily infer that the face he makes, the jerking movement of his body, and the tortured utterance, "Ouch!" must mean he is in pain.

In some other cases, what one perceives and the reasons one attributes for what one perceives are quite different from one another. Suppose that friend of mine has promised me three times that we will go out to have a drink. The initiative, however, needs always to come from me. This situation is more complicated. It could be that my friend really has no time, forgets our appointments, dislikes me but is embarrassed to say so, a combination of these factors, etc. In such cases, more elaborate narratives can be used as "heuristic fictions" to join all the clues – both cognitive and affective – into a reasonable explanation. In turn, this explanation has a cognitive-affective impact. When I – after due deliberation – conclude that my (so-called) friend does not like my company, this realization may go hand in hand with feelings of anger, frustration, or disappointment.

That infants can understand facial expressions, gestures and body language at large is clear. Arguably, they do not create narratives or engage in scenario-thinking. While they understand bodily *expressions* of emotions, I think it is far-fetched to say that they understand (always) *why* these emotions are being displayed.

---

[72] Gallagher (2011: 58).
[73] Gallagher (2009: 253).

Understanding *expression* does not equate with understanding *intention*. However, precisely that is the domain of ST. The central claim of ST (at least as I would construe it) is not to be a theory about our primary mode of understanding. Instead, I believe it to be a predictive strategy with which human beings capable of doing so generate explanations, predictions, narratives and fictional options. In turn, the products of these simulative efforts are not merely ratiocinations about behaviour, but they consist of cognitive-affective states that cannot be decoupled from our embodiment in the world.

Symbolic interaction is already present in pointing and cooperative action, but reaches its crucial stage in language. Verbal narratives then become the presupposition for more sophisticated modes of understanding which develop in the third and fourth year of life. By engaging in storytelling practices, children learn to understand others in a meaningful way, to imagine their goals and intentions as underlying a certain course of actions.[74]

Admittedly, both TT and ST have a bit too often been presented as comprehensive Theories of Mind, or in other words, as theoretical accounts that could reduce a vast array of human interactions to a single, universally valid principle, whether this was theorizing, simulating, or a mix of both. In this connection, Daly is spot-on correct when she says that "simulation [theory] was too mentalistic."[75] As discussed in the first sections of this article, the conception of mental simulation as a purely "boxological,"procedural affair passes too lightly over the embodied and affective components of everyday social cognition, and the irreducibly cognitive-affective character of the imagination. The reduction of an interpretive strategy so dependent on affect to either logical theorizing (as in TT) or pure heuristics (as in ST) discards embodiment and presents a rather impoverished version of intersubjectivity.

In this regard, I largely agree with Gallagher when he states that TT and ST in general share these presuppositions:[76]

> (A) The problem of social cognition is due to the lack of access that we have to the other person's mental states. Since we cannot directly perceive the other's thoughts, feelings, or intentions, we need some extra-perceptual cognitive process (mindreading or mentalizing) that will allow us to infer or simulate what they are.

---

[74] Fuchs (2015: 194).
[75] Daly (2014: 231–232).
[76] Gallagher (2011: 56–57).

(B) Our normal everyday stance toward the other person is a third-person, observational stance. Based on what we observe we use mindreading to explain or predict their behaviors.

(C) These mentalizing processes constitute our primary and pervasive way of understanding others

The problem of social cognition (aka "the problem of other minds") is due to the lack of access that we have to the other person's mental states. Since we cannot directly perceive the other's thoughts, feelings, or intentions, we need some extra-perceptual cognitive process (mindreading or mentalizing) that will allow us to infer or simulate what they are. Our normal everyday stance toward the other person is a third-person, observational stance. Based on what we observe we use mindreading to explain or predict their behaviors. These mentalizing processes constitute our primary and pervasive way of understanding others.

Gallagher paints (understandably) with broad strokes here, and aims primarily at older accounts of ST. Therefore, some of the distinctions he introduces here do not touch on imagination-driven simulation.

For instance, imagination-driven simulation as presented here partially disagrees with (A). The idea that we need an "extra-perceptual process" seems to rest on the idea that a neat division can be made between perceptions (via the senses) and mental processes dealing with them. However, imagination-driven simulation emphasizes that mental states are affective-cognitive. The perception and processing of what is perceived are both accomplished in a fully embodied manner. This is not to say that some mental processes and neurobiological phenomena (such as the presence of mirror neurons) cannot be identified or studied individually. Instead, it means that the claims that the isolated study of such processes can have on imagination-driven simulation are limited. One cannot easily base a Theory of Mind on purely empirical findings, and directly extrapolate from findings in, for example, neuroscience, to a comprehensive Theory of Mind.

Imagination-driven simulation partially disagrees with (B) because imagination-driven simulation is not reducible to (especially Goldman's account of) mindreading. Some mindreading is involved in the simulative process, and it arguably plays an important role. To hold that therefore mental simulation is reducible to mindreading is to drive the reduction too far, confining simulation again to brain activity.

Likewise, I disagree with (C) that merely "mentalizing processes" constitute our primary way of understanding others. If we would concede this, we would be

back at a kind of dualism, wherein all activity that matters for understanding others is located in the brain, and the body is merely a mute device for keeping the brain alive. Gallagher's ironic phantasy of the both TT and ST theorists as "body snatchers" seems spot-on to me on this point. Often, older TT and ST theorists pay lip service to the idea of embodied cognition, just in order to maintain that the brain is the "place where everything *really* happens."[77]

Gallagher's own Interaction Theory (IT) postulates three different presuppositions that are aimed at negating those underpinning both TT and ST, of which I will discuss two here:[78]

> (D) Other minds are not hidden away and inaccessible. The other person's intentions, emotions, and dispositions are expressed in their embodied behavior. In most cases of everyday interaction no inference or projection to mental states beyond those expressions and behaviors is necessary.
>
> (E) Our normal everyday stance toward the other person is not third-person, detached observation; it is second-person interaction. We are not primarily spectators or observers of other people's actions; for the most part we are interacting with them in some communicative action, on some project, in some pre-defined relation; or we are treating them as potential interactors.

I have already discussed point (D) to some degree. Simulation takes place when the gap between that which is perceived and the reasons for this action becomes obvious. When I observe unexplainable or curious behaviour, or when I find myself in an unfamiliar situation, imagination-driven simulation is a strategy to help me cope with my lack of information.

One thing needs to be said about the "third-person account" that Gallagher criticizes in point (E). According to him, we are routinely caught up in a "second-person account,"since we deal continuously with one another, making social cognition a skill that is learned through immersion rather than observation. Moreover, it is partially defined by social relations. Gallagher is certainly right that we do not move about in the world while consciously simulating our environment, inferring clues, and relentlessly producing conjectures. That would be a horrible and unimaginably tiring prospect. The point where one could disagree with Gallagher is his assertion that "we are not primarily spectators or observers of other people's actions." It seems to me, on the contrary, that in an important way, we *are* primarily spectators or observers of the Other. As he later on contends in the same paper, we

---

[77] See Gallagher (2015).
[78] Gallagher (2011: 59).

are "thrown into" the world, without an option to distance ourselves. The decision is already taken for us.

Otherwise put, short of Robinson Crusoe scenarios (and of course Crusoe himself finally met Friday) we cannot help but be spectators or observers of others, dealing with them as our embodiment dictates. As discussed, Sartre is here right when he states that we "constitute others" along certain lines. The line that Sartre blurs here is between observation and immersion. Observation is on this account already (implicit) judgement and affect-laden. There is no such thing as a "disembodied observation." Conversely, Gallagher is certainly right that some early work on ST oriented itself too much towards observing, implicitly assuming that we are caught up in mental processes of observation-and-inference. But Gallagher's own account of second-person interaction leans heavily on a hard split between observation and immersion. Even if it is true that our day-to-day encounters take place from a "second-person point of view," the observational stance that Gallagher rejects does not always play the role he seems to assume. Of course, there are moments where we consciously observe others (not coincidentally, watching people is a popular pastime activity). However, in more mundane cases, the observational is "included in the glance," as it were. As Gallagher himself points out, we cannot help seeing the world as immersed observers. The fact that we are irreducibly immersed and embodied furnishes us with a constitutive glance insofar as we are able to form concepts.

Imagination-driven simulation and IT have many ideas in common. Of course, one could disagree about details, but it seems to me that Gallagher's account could be frutifully enriched with the kind of modified simulation theory I have presented here.[79] Conversely, IT seems a promising venue for staging our departure from the overtly mentalistic, older conceptions of mental simulation.

[79] Consequently, I do not share Gallagher's worry that elements of IT may be too easily be appropriated by TT and ST (Gallagher [2009: 295]). This looks like a desirable scenario to me, where decades of accumulated reasoning result in integrated, well-balanced theories.

REFERENCES

Barz, W. 2014. "Introspection as a Game of Make-Believe." *Theoria* 80: 350-367

Bermúdez, J. and Cahen, A. 2015. "Nonconceptual Mental Content." In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy (Fall 2015 Edition)*. Available online at URL = <https://plato.stanford.edu/archives/fall2015/entries/content-nonconceptual/>.

Carruthers, P. 1998. "Simulation and Self-Knowledge: a Defence of Theory-Theory." In P. Carruthers and P.K. Smith (eds.), *Theories of Theories of Mind*. 2nd edn. Cambridge, Cambridge University Press. 22-38

Churchland, P. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78 (1981): 67-90.

Currie, G. 1995. "Imagination and Simulation: Aesthetics meets Cognitive Science." In M. Davies and T. Stone (eds.), *Mental Simulation: Evaluations and Applications — Reading in Mind and Language*. Oxford: Wiley-Blackwell. 151-169.

Daly, A., 2014. "Primary Intersubjectivity: Empathy, Affective Reversibility, 'Self-Affection' and the Primordial 'We'." *Topoi* 33: 227–241.

Davies, M. 1998. "Simulation Theory." In *Routledge Encyclopedia of Philosophy*. Available online at URL = <https://www-rep-routledge-om.libezproxy.open.ac.uk/articles/thematic/simulation-theory/v-1/sections/simulation-in-reality-and-in-imagination>.

Dennett, D. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Fuchs, T. 2015. "Pathologies of Intersubjectivity in Autism and Schizophrenia." *Journal of Consciousness Studies* 22: 191–214.

Gallagher, S. 2015. "Invasion of the Body Snatchers: How Embodied Cognition is Being Disembodied." *The Philosophers' Magazine*. 96-102.

------. 2011. "Strong Interaction and Self-Agency." *Humana.Mente* 15: 55–76.

------. 2009. "Two Problems of Intersubjectivity." *Journal of Consciousness Studies* 16: 289–308.

Gallese, V., and Goldman, A. 1998. "Mirror Neurons and the Simulation Theory of Mind-Reading." *Trends in Cognitive Science* 2: 493-501.

Garrett, D. 2005. "The Contents and Faculties of the Mind." *Routledge Encyclopedia of Philosophy*, 2005. Available online at URL = <https://www-rep-routledge-com.libezproxy.open.ac.uk/articles/biographical/hume-david-1711-76/v-2/sections/the-contents-and-faculties-of-the-mind>.

Goldman, A. 2009. "Mirroring, Simulating and Mindreading." *Mind and Language* 24: 235-252.

------. 1995. "Interpretation Psychologized." In M. Davies and T. Stone (eds.), *Folk Psychology: The Theory of Mind Debate*. Oxford: Wiley-Blackwell. 74-99. Originally published in *Mind and Language* 4 (1989): 161-185

Gordon, R.M. 2007. "Ascent Routines for Propositional Attitudes."*Synthese* 59: 151-165.

------. 2001. "Simulation and Reason Explanation: The Radical View." *Philosophical Topics* 29: 175-192

------. 1998. " 'Radical' Simulation." In Carruthers and Smith (eds.), *Theories of Theories of Mind*. 11-21.

------. 1995b. "The Simulation Theory: Objections and Misconceptions." In M. Davies and T. Stone (eds.) *Folk Psychology: The Theory of Mind Debate*. Oxford: Wiley-Blackwell. 100-122. Originally published in *Mind and Language* 7 (1992): 11-34.

------. 1995a. "Folk Psychology as Simulation." In Davies and Stone (eds.), *Folk Psychology: The Theory of Mind Debate*. 60-73. Originally published in *Mind and Language* 1 (1986): 158-171.

Heal, J. 2000. "Other Minds, Rationality and Analogy." *Proceedings of the Aristotelian Society* 74: 1-19.

------. 1998. "Co-Cognition and Off-Line Simulation: Two Ways of Understanding the Simulation Approach." *Mind and Language* 13: 477–498.

------. 1995. "Replication and Functionalism." In Davies and Stone (eds.) *Folk Psychology: The Theory of Mind Debate*. 45-59. Originally published in Butterfield, J., (ed.), *Language, Mind, and Logic*. Cambridge: Cambridge University Press, 1986. 135-150.

Hume, D. 1777. *An Enquiry Concerning Human Understanding*. Available online at URL = <http://davidhume.org./texts/ehu.html>.

------. 1740. *Abstract of a Book lately Published.* Available online at URL = <http://davidhume.org./texts/thn.html>.

------. 1739. *A Treatise of Human Nature*. Available online at URL = <http://davidhume.org./texts/thn.html>.

Kant, I. 2009. *Critique of Pure Reason*. Trans. P. Guyer and A. Wood. Cambridge: Cambridge University Press.

Leslie, A.M., and German, T. 1995. "Knowledge and Ability in 'Theory of Mind': One-eyed Overview of a Debate." In Davies and Stone (eds.) *Mental Simulation: Evaluations and Applications—Reading in Mind and Language*. 123-150.

Machery, E. 2005. "You Don't Know How You Think: Introspection and the Language of Thought." *British Journal for the Philosophy of Science* 56: 469-485.

McGinn, C. 2009. "Imagination." In B. McLaughlin, A. Beckermann, and S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford Univ. Press. 595–606.

Nichols, S., Stich, S., Leslie, A. and Klein, D. 1998. "Varieties of Off-Line Simulation." In Carruthers and Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge Univ. 39-74.

Nisbett, R.E. and Wilson, T. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (1977): 231-259

Proust, J. 2002. "Can Radical Simulation Theories Explain Psychological Concept Acquisition?" In J. Dokic and J. Proust (eds.), *Simulation and Knowledge of Action*. Amsterdam: John Benjamins. 201-228.

Sartre, J. P. 2004. *The Imaginary: A Phenomenological Psychology of the Imagination*. Trans. J. Webber. London: Routledge. Originally published by Éditions Gallimard, Paris, 1940.

Saxe, R. 2005. "Against Simulation: The Argument from Error." *Trends in Cognitive Science* 9: 174-179.

Stich, S. and Nichols, S. 1998. "Theory Theory to the Max." *Mind and Language* 13: 421-449.

------. 1997. "Cognitive Penetrability, Rationality and Restricted Simulation." *Mind and Language* 12: 297-326.

------. 1995. "Second Thoughts on Simulation." In Davies and Stone (eds.), *Mental Simulation: Evaluations and Applications—Reading in Mind and Language*. 87-109.

Wittgenstein, L. 2009. *Philosophical Investigations*. 4th edition. Trans. P.M.S. Hacker and J. Schulte. London: Wiley-Blackwell. Originally published by Basil Blackwell Ltd., London, 1953.

------. 1980. *Remarks on the Philosophy of Psychology*. Trans. G.E.M. Anscombe. 2 vols. Chicago IL: Univ. of Chicago Press.