Croatian Journal of Philosophy
Vol. XVII, No. 52, 2018

Simulation and Thought Experiments. The Example of Contractualism

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
Central European University, Budapest, Hungary

The paper investigates some mechanisms of thought-experimenting, and explores the role of perspective taking, in particular of mental simulation, in political thought-experiments, focusing for the most part on contractualist ones. It thus brings together two blossoming traditions: the study of perspective taking and methodology of thought-experiments. How do contractualist thought-experiments work? Our moderately inflationist mental modelling proposal is that they mobilize our imaginative capacity for perspective taking, most probably perspective taking through simulation. The framework suggests the answers to questions that are often raised for other kinds of thought-experiments as well, concerning their source of data, heuristic superiority to deduction, experiential, qualitative character and ease in eliminating alternatives. In the case of contractualist political thought-experiments, the data come from perspective taking and the capacity to simulate. Mental simulation is way more accessible to subjects than abstract political reasoning from principles and facts. There is a new experience for the subject, the one of simulating. Simulation normally is quick and effortless; the simulator does not go through alternatives, but is constrained in an unconscious way. We distinguish two kinds of political thought-experiments and two manners of imagining political arrangements, building third-person mental models, and first-person perspective taking. The two mechanisms, the first of inductive model building, the second for simulation, and their combination(s), exhaust the range of cognitive mechanism underlying political thought-experimenting.

Keywords: Thought experiment, simulation, social contract, veil of ignorance.

1. *Introduction*

A lot of thought experiments (TEs) requires the reader to take perspective on some morally, politically or legally relevant imagined situation; the Golden Rule TEs normally require one to take perspective on the victim's situation, the Veil-of-ignorance TE to take perspective on possible social arrangements under the supposition that one is ignorant of her own material situation, abilities and the like.[1] The topic of perspective-taking has become extremely popular in philosophy, psychology and related disciplines, in particular as far as its empathetic version is concerned.[2] In this paper I shall explore the role of perspective taking in political TEs (for short "PTEs"). What is the actual cognitive mechanism underlying the process? Here I shall opt for one particular, and rather popular view on perspective taking, namely that it crucially involves mental simulation (see Goldman 2006). Goldman has, of course noticed, the connection to various TEs, in particular to Golden Rule and the Veil-of-ignorance ones (2006: 294), but has not been developing it much. So, the goal here is accounting for cognitive mechanism underlying political thought-experiments (PTEs), more narrowly upon the presently most popular variant, namely the contractualist ones, in the widest sense of the term, with authors like Rawls, Scanlon, Habermas and Parfit (see References) at the forefront. These experiments typically address any given issue about the moral and political status of some arrangement (say, the status of the right to privacy) by inviting the reader to imagine a situation in which she is enabled to choose in the favor of it or against it, in her own name, and/or in the name of other people, under specified circumstances. She might be asked to imagine having to persuade other people to accept her choice, and reflect about ways of doing it, and so on. At the end of the experiment, the reader is supposed to have arrived at intuition(s) concerning the issue, for instance that she would choose the arrangement under such-and-such circumstances (say, under the Veil-of-ignorance), or that most people could not be persuaded to accept it, again under specified circumstances. These intuitions are not themselves normative, they are factual intuitions about possible choices. However, they serve as the basis for further theory-building, which then results in normative conclusions, usually of moral-cum-political character. The question this paper is addressing is simple to state: *Where do the intuitions come from? What is the possible psychological mechanism that produces the factual intuitions that serve as the basis for normative theory*?

The framework for the answer shall be my moderately optimistic, "deflationist" as David Davies (2018) calls it, mental modelling

[1] The paper originated from a presentation at a conference in Geneva 8–9 June 2017, on "Simulation and thought experiment". I would like to thank prof. Marcel Weber for inviting me, and the participants for interesting and helpful discussion.

[2] See chapters in Coplan and Goldie (2011) and Maibom (2017).

approach.[3] I agree with him about the characterization, and I thank him. A variant of it has been developed in detail, namely the one that concerns building a mental model from the third-person perspective. I hope it can account for PTEs like Plato's *Republic*, where a group of young elite Athenians is supposed to imagine what life would be like for all sorts of people in a philosophers ruled state (Miščević 2012a) In general, it is suitable for imagining political arrangements, primarily from the third-person perspective. However, it does leave open the accounting for a different kind of modelling, in which the experimenter is imagining social-political situations and arrangements, primarily from the first—person perspective—the social contract (SC) tradition and its present-day form, with star author like Rawls, Scanlon, and Habermas. For this kind of thought experiments I want to propose a solution within the general framework of mental modelling, but stressing a different kind of it: not building a model from the third person perspective, but trying to imagine how things would look to oneself, from the first person perspective. I will opt for one theory of such enactive imagining, namely the idea that we simulate perspective taking.[4] Here is then the preview.

Section 2.1 summarizes the main idea of the SC tradition, also mentioning a simple forerunner of SC idea, namely the Golden Rule proposal. It proposes a division of SC theories, contrasting first the hypothetical ones, and the non-hypothetical (or partly non-hypothetical ones), and then, within the first group, those that rely on the picture of real, "normal" contractors, and those that propose idealization or other kinds of "retouch" of the parties participating. The SC PTE is built around the question for the would-be participants: what kind of arrangement would you accept, find just and liveable? The subject is supposed to arrive at an intuitional answer to the question.

Section 2.2 is the central part of the paper, dedicated to accounting for PTEs, in particular for the epistemic-psychological side and the question of how the relevant intuition gets formed. The first, very brief, subsection concerns the structure of a TE, and the second one turns to the role of simulation, that will be presented as the royal road to intuition. After a general brief mention of theories of simulation, it turns to its role in practical TEs. This is the central sub-section and the most important part of the paper, stressing the central role of simulation and showing how it fits well with independently established requirements of contractualist PTEs.

Section 2.3 mentions some difficulties with simulation that have been pointed to in the literature and offers some optimistic answer to them. In Conclusion we briefly sketch the bigger picture, namely our

[3] I started defending this approach quarter a century ago, or, to put it more accurately a variant of it, in Miščević (1992).

[4] Of course, some kind of first person imagining might have been required in the *Republic* scenario: how would you feel if you had to lead the polis, and so on, but it is not central, as it will become in the SC tradition.

proposal for accounting for cognitive mechanism underlying PTEs in general, hoping that it might help understanding thought-experimenting in general and thus throw an additional light upon the foundations of methodology of philosophy.

Let me in the rest of this section introduce the kind of PTEs we shall be dealing with, namely the works in Social contract (SC) tradition that invite the reader to imagine social-political situations and arrangements that can come from the willingness of parties to come together and negotiate the best possibility. The first modern authors, Hobbes and Locke, are not completely clear about the factual status of the Contract, whether it is a historical event or merely imagined one; with Rousseau and most explicitly Kant, it becomes "hypothetical" contract, fit to be classified as a TE. At least since Rousseau it is discussed primarily from the first-person perspective; the typical leading question is Would *you* sign a contract ...under such-and-such conditions? We shall here set on one side contractarian version (due to authors like Hobbes and Gauthier (1986)) focused on maximization of the self-interest of each participant, since it poses less challenging problems to participant's imagination, and focus on the more challenging contractualist line with authors like Rawls, Scanlon, Habermas and Parfit (see References). The typical demand here is to put oneself in another's shoes while asking yourself: what demands cannot be rejected by my interlocutor, if she is rational?

We can describe the crucial imaginative exercises as moral and political TEs from the first-person perspective. This makes the SC tradition contrast with another famous tradition of thought-experimenting, starting at least with Plato's *Republic*: building and understanding a complex social arrangement primarily from the third-person perspective (tell me, Glaucon, how would you judge the commonality of children? is it just or not? and so on...). It continues by building a mega-arrangement, and in more practically oriented cases ends as a utopia or dystopia, so to speak, with famous authors like Al Farabi, Thomas More and Fourrier. Let me mention a contemporary proposal from the third-person perspective, a simple and fine example: the camping trip and equality among campers in G. A. Cohen *Why Not Socialism*. On camping people exercise solidarity, treat each other as equal, help altruistically and without reservations, so, Cohen concludes, we can use it as a model for a socialist society Cohen uses the understanding of equality provided by the trip-model to argue for extremely high level of equality in his socialistic society. (Some cases that are difficult to classify, say, prominently Dworkin's anti-luck TEs).

Back to contractualism. Let me quickly propose a systematization of the main philosophical proposals within the hypothetical contract views since they are most relevant for discussion of thought-experimental methodologies, all this with apologies for brevity. One line does not propose, at least explicitly, any retouch of ordinary circumstances: the participants are real people, endowed just with ordinary rationality. Kant

and Parfit (in *On What Matters*) are prime examples of such approach.

With Rawls, a different line of experimenting started. The real subjects are replaced or supplemented by somewhat "retouched" model participants; in Rawls' work the "parties" in the Original position, are famously placed behind the Veil-of-ignorance, and they just "represent" the real persons who make their contract on the basis of principles figured out by the "parties".[5] In the Original position the person decides to try the Veil-of-ignorance; she attempts to answer the crucial question: what arrangement would you choose if you were ignorant of some important aspects of your future situation? You ask yourself: shall I be male? Or female? And what is the best decision to take if I don't know the answer? Shall I be intelligent? Or stupid? And so on.

Her counterpart, the "party" behind the Veil has to do the job:

> The idea here is simply to make vivid to ourselves the restrictions that it seems reasonable to impose on arguments for principles of justice, and therefore on these principles themselves. Thus it seems reasonable and generally acceptable that no one should be advantaged or disadvantaged by natural fortune or social circumstances in the choice of principles. It also seems widely agreed that it should be impossible to tailor principles to the circumstances of one's own case. We should insure further that particular inclinations and aspirations, and persons' conceptions of their good do not affect the principles adopted. (Rawls 1999: 16)

The typical questions concern wealth, status, talents and the like. How would you decide if you knew you will be poor? Or, deprived of interesting and important talents? The person's identity is preserved, and she simulates her reaction in a different situation than her actual one. The reason why in the Original position she has to deprive the participants of concrete knowledge of their actual standing in various relations in society is that participants have working models of social interaction. Therefore, if the person is choosing rationally, she will be partial to his (actual and future) self, and the promise of justice will be gone. Now, behind the Veil the participant does not know how rich she will be. She has to imagine herself being very rich (vow!), being moderately well off (not bad!) and being very poor (God forbid!).

Since parties have rich general information, she uses her default knowledge of how it feels being rich, well off and poor. She does not proceed to building a further model from the third-person perspective, but reasoning from the first-person perspective: let me imagine myself being poor, etc.! It is here that we shall introduce the idea of simulation. And the imagining will result in producing an answer, a particu-

[5] Rawls in his *The Basic Liberties and Their Priority, The Tanner Lectures on Human Values April 10, 1981* stresses the following advice: "Two different parts of the original position must be carefully distinguished. These parts correspond to the two powers of moral personality, or to what I have called the capacity to be reasonable and the capacity to be rational. While the original position as a whole represents both moral powers, and therefore represents the full conception of the person, the parties as rationally autonomous representatives of persons in society represent only the Rational (...)." In McMurrin (1986: 19).

lar intuition: I should secure myself against the risks of ending up in a very bad situation.

Let me just mention the other famous retouch options. The main alternative to ignorance is idealization: how would my interlocutor react if she were made a bit more rational, and the discussion and decision situation were closer to an ideal one? Again, we are invited to put ourselves in another's shoes, this time in the shoes of a richly rational person, in the sense of rationality that also includes moral sensibility (in contrast to the means-end rationality of parties behind Rawls' Veil. What demands cannot be rejected by my interlocutor, if she is rational, Thomas Scanlon is asking:

> My view ... holds that thinking about right and wrong is, at the most basic level, thinking about what could be justified to others on grounds that they, if appropriately motivated, could not reasonably reject. On this view the idea of justifiability to others is taken to be basic in two ways. First, it is by thinking about what could be justified to others on grounds that they could not reasonably reject that we determine the shape of more specific moral notions such as murder or betrayal. Second, the idea that we have reason to avoid actions that could not be justified in this way accounts for the distinctive normative force of moral wrongness. (Scanlon 1998 :5)

The procedure is presented as valid for political, institutional arrangements as well as for individual morality. Scanlon talks about "standards that institutions must meet if they are to be justifiable to those to whom they claim to apply" (Scanlon 2016: 5). So, suppose I want to propose a practice or institution *P* that is to apply to you. I have to get into your shoes: what kind of arguments would rationally persuade you to accept *P*? And a particular intuitional answer follows.

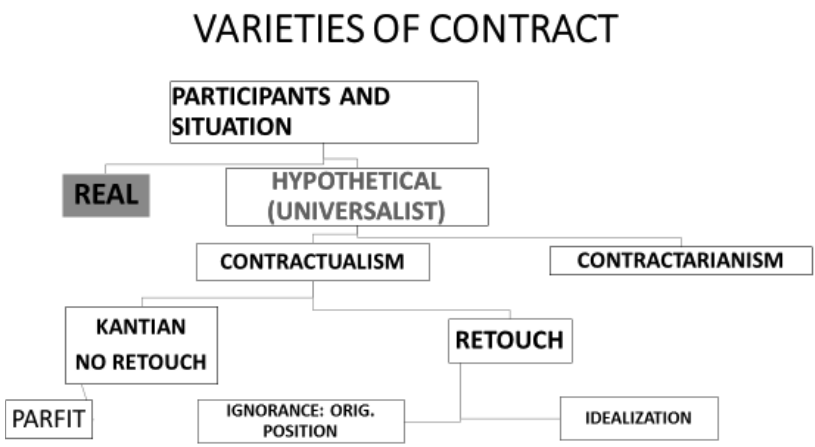
Finally, let me mention Habermas, who explicitly talks about his proposal as a TE (1989), and proposes to introduce idealized communicative situation as a whole, not just idealizations concerning the participants. Here is a brief characterization from the chapter "Remarks on Discourse Ethics":

> The notion that ideal role taking-that is, checking and reciprocally reversing interpretive perspectives under the general communicative presuppositions of the practice of argumentation-becomes both possible and necessary loses its strangeness when we reflect that the principle of universalization merely makes explicit what it means for a norm to be able to claim validity. Already in Kant the moral principle is designed to explicate the meaning of the validity of norms; it expresses, with specific reference to normative propositions, the *general* intuition that true or correct statements are not valid just for you or me alone. Valid statements must admit of justification by appeal to reasons that could convince anyone irrespective of time or place. In raising claims to validity, speakers and hearers transcend the provincial standards of a merely particular community of interpreters and their spatiotemporally localized communicative practice. (Habermas 1994: 52).

We are invited to see idealizations as those simultaneously unavoidable and trivial accomplishments that sustain communicative action and argumentation (Habremas 1994: 55). Commonsensical moves, like

attributing identical meanings to expressions, attaching "context-transcending significance to validity claims", and ascription of rationality and accountability to speakers are pragmatic presupposition of communication that involve some idealization. The philosophical idealizations just continue where the ordinary ones stop.

In short, we thus have "retouched" (distorted) model participants and situations, changed basically in two directions. First, ignorance: what arrangement would you choose if you were ignorant of some important aspects of your future situation? Second, idealization what demands cannot be rejected by my interlocutor, if she is rational? And if we are placed in an ideal communicative situation? And here is the scheme of the division:



Of course, the proponents of the ignorance strategy, Rawls and his many followers, have been confronting the defenders of idealization, like Habermas and Scanlon with their followers, and vice versa, and the debate has reached epic proportions. However, we have to leave it for another occasion, and pass to our main topic, the mechanism that produces the answers-intuitions.

2. Accounting for PTEs: The Role of Simulation

2.1 The task ahead

How do people understand imaginative scenarios essential for PTEs? Not much has been written about mechanism underlying PTEs. We need a more detailed look at TEs in general, and PTEs in particular.

Take first the simplest example, the Golden rule. Suppose I am bragging around with my knowledge of some area, and letting my colleague know how ignorant and incapable they are, when it comes to important issues. My wife asks: "Well, how would you feel if somebody were doing this to you?" I am supposed to imagine the reversed situation, go through the process of being humiliated, and feel what people normally

feel in such situations. This should make me sensitive to what I am doing. Here is Parfit's description of the typical process:

> When we apply the Golden Rule, our thought-experiment is fairly simple. As when making many ordinary decisions, we ask what would happen in the actual world if we acted, on one occasion, in each of certain possible ways.
>
> We don't even need to decide what are the morally relevant descriptions of these particular possible acts. But we try to think about these possibilities, not only from our own point of view, but also from the points of view of all of the other people whom our act might affect. We ask what we would rationally be willing to do, and have done to us, if we were going to be in all of these people's positions, and would be relevantly like them. (Parfit 2011: 328)

Let me propose a picture of the process of reasoning in a TE. We have two persons, the experimenter and the subject. At the preliminary stage, call it *stage 0*, the experimenter formulation her design: in our example, show to me that I should not humiliate my colleagues, and she wants to do this by asking me to imagine switching the role with a colleague, call him Jack.

At *stage one*, comes the presentation of the scenario thus constructed to the experimental subject, in this example to me: imagine you are Jack, the person that you have been humiliating!

At *stage two,* I, the experimental subject, come to understand the question. For instance, how would you feel if somebody were doing this to you?

At *stage three* comes the tentative production, "modeling" of the scenario at the conscious level. I imagine being humiliated. Then some unconscious processing might get in. The stage concerns the production of the answer, involving the generation of intuition; for instance, how I would feel in the shoes of the victim. This probably involves reasoning at the unconscious level; for instance, I might have to control my arrogance, and belief that yes, my colleagues are not as good as I am, and the like. This might result in an immediate, unconscious intuition, e.g. Yes, I would feel terribly...

At the next, *fourth stage*, the thinker comes out with explicit intuition at the conscious level, usually geared to the particular example and having little generality (again, I would feel terribly, etc.). This ends the core TE

Usually however, there is a *fifth* stage. The thinker often has to do some varying and generalizing, at the conscious and reflective level and, perhaps, at the unconscious one too. For instance, in the story I might be unimpressed by threat concerning my professional abilities. Imagine then, my wife might say, your young colleagues making deprecatory remarks about your age, suggesting it's time for you to retire, and let more energetic, younger people occupy the stage. And imagine that this is done by a younger colleague, what if it is done by a brilliant doctorate student, of someone else, over whom I have no power? Sometimes this process of going through related micro-TEs is called intui-

tive induction (Chisholm 1966). I end up with a general belief that my behavior is morally not acceptable. No matter what, such *kind of treatment* is awful, I would feel this for sure if someone did treat me thus.

If I am reflective enough, I might go one step further, to stage *six*. I first, consciously perform the aggregation of micro-TEs; second, I try to harmonize the result s of these micro-TEs with each other, and finally, I arrive at a judgment of their coherence with other moral intuitions one might. In other words, this philosophical unification can be described in terms of reflective equilibrium, first narrow and then wide. Here, the general knowledge of more empirical kind is brought into play. I arrive at important and difficult task of comparing the result with all we know about life and politics, both personal experiential level and from history, social and natural sciences, reaching a wide reflective equilibrium as the final result.

A similar, but more demanding process goes on in the case of a contractualist TE. Take the Veil-of-ignorance situation and assume you are a male. Now you are asked to ponder the following Rawlsian question: what distributive arrangement would I chose if I didn't know whether I will be rich or poor? I basically go through same or analogous stages, and reach the final (non-moral) intuition, say I don't want to risk extreme forms of poverty, I want a decent life even if I am not rich. (Habermas similarly talks about "interlocking of perspectives", where everyone is required to take the perspective of "everyone else (1995: 117)).

Here, we shall be mostly interested in stages three and four where this is supposed to occur. How does the thinker model the situation proposed in the scenario, and how is the resulting intuition produced? For this, we turn to cognitive investigations.

2.2. *Simulation, the royal road to intuition*

We have implicitly pointed to a promising answer: the thinker arrives to her intuition through mental modelling. I have been defending the role of mental modelling for more than two and a half decades (see Miščević 1992). David Davies mentions that I set "out clearly (1992: 24) how this approach solves the usual puzzles about TEs" (2018: 520). TEs enable us to produce new data by manipulating old data through the generation of a manipulable representation of a problem. In constructing and manipulating this model, we mobilize prior cognitive resources in new ways.

I would go further and claim that what cognitive science tells us about perspective taking, and more particularly about simulation, offers an interesting variety of this answer to our question. The idea that simulation produces the relevant intuition suggests the role of competences in TEs. I have been conjecturing that some of them might be quite general (the capacity to simulate other person's mental states which will occupy us in the sequel), some less general (folk-physics),

and some completely specialized (spatial, linguistic and mathematical skills). This suggests that the typical verdicts from TEs are in a way voice of competence, albeit a discreet one, often mixed with those from other sources (general intelligence, social skills, emotional life) (see Miščević 2006, 2012). Here I want to introduce some new proposals, focusing on our capacity for perspective taking.

Let me distinguish two kinds of mental modelling that often get confused in the literature. One is the *third-person model-building*, for instance imagining a planet when reading a science-fiction story or Putnam's Twin earth description. The planet is an object that is imagined from a third-person perspective, relatively static, although allowing for some imagined movement. The other is the kind that will interest us here: *the first-person process of modelling, typically through mental simulation*, in which the subject imagines herself as the protagonist.

Let me first say a few words about the third-person model-building. Mental models, psychologists tell us, purport to represent concrete situations, with determinate objects and relations (precisely what is demanded in thought experiments) (Johnson-Laird 1983: 157). Their structure is not arbitrary, but "plays a direct representational role since it is analogous to the corresponding state of affairs in the world" (Johnson-Laird 1983: 157). One can distinguish simple static "frames" representing relations between a set of objects, crucially of human beings in the PTEs, temporal models consisting of sequences of such frames, kinematic models which is the temporal model with continuous time, and finally dynamic models which model causal relations. Reasoning in mental models demands rules for manipulation. Johnson-Laird hypothesizes the existence of general procedures which add new elements to the model, and 'a procedure that integrates two or more hitherto separated models if an assertion interrelates entities in them'. The integration of models is subject to consistency requirements—if the joint model is logically impossible, some change has to be made (Miščević 1992: 220).[6]

Can this model-building from the third-person perspective help us with examples like Golden Rule and Veil-of-ignorance? Doesn't look very promising. The sinner in the Golden Rule experimenting is not supposed to imagine a neutral, distanced situation; she has to imagine *herself* in the reversed situation. Similarly, the thinker behind the Veil asks herself how *she herself* would choose, from the first-person perspective.

Here, a plausible alternative mechanism that would enable modelling from the first-person perspective is a mechanism of perspective taking. Psychologists talk about various ways of simulating and have interesting things to say about this. Consider the famous psychologists C. Daniel Batson. In his (2009) paper he writes:

> Encounter a stranger in need and, sometimes, you will feel empathic con-

[6] Other authors in the similar vein are Zwaan and Radvansky (1998), Hohwy (2013), and Frith (2007).

> cern—an other-oriented emotional response evoked by and congruent with the perceived welfare of that person. What determines whether you will? Perhaps the most common answer among psychologists is that empathic concern is felt when you adopt the perspective of the person in need (…). But in this answer, what is meant by adopting the person's perspective? First, it is an act of imagination. One does not literally take another person's place or look through his or her eyes. One imagines how things look from the other's point of view. Second, it is not the same as perspective taking in the symbolic-interaction tradition (…). In that tradition, one adopts the perspective of another—often a significant other—to imaginatively see oneself through the other's eyes (and values). The perspective taking that evokes empathic concern involves imaginatively perceiving the other's situation, not oneself. (Batson 2009: 267)

Philosophers and psychologists talk a lot about empathy, as a paradigmatic kind of perspective taking. However, there are several terminological problem connected with the term "empathy"; first, often carries connotations of "sympathy", so that empathetic understanding is the one that is accompanied by sympathetic feelings.[7] Famous authors routinely connect the two (de Waal 2009, Clohesy 2013).[8]

Since we find mental Simulation Theory the best account of perspective taking, we shall turn to it and talk about "mental simulation" as the relevant activity. We hope, however, that what we have to say holds for perspective taking in general, and, if Simulation Theory turns out to be defective, can be connected to whatever account of perspective taking replaces it. We shall thus speak of perspective taking, and in particular of mental simulation as the second kind of modelling, besides the third-person one, that we need in order to answer the question of how we arrive at our responses and other people's ones in cases like Golden Rule or the Veil. Indeed, some authors relying on cognitive science count ability to simulate as a part of general modelling ability. Thomas Metzinger mentions important traits of mental models, like being multimodal, mutually embeddable, often analog rather than digital, and then adds *ability to simulate (independently from input)* (2003: 109ff.)

So, let us turn to simulation. We are mental simulators, not in the sense that we merely simulate mentation, but in the sense that we understand others by using our own mentation in a process of simulation,

[7] For relevant warnings see Amy Coplan (2011: 3). We shall heed the warnings and avoid unqualified use of the term.

[8] Here is a statement by psychologist Chris Frith: "One obvious question is why have we put together empathy and fairness? In neuroscience there is not much overlap in the literature on these topics. Fairness tends to be studied within the realm of neuroeconomics, whereas empathy springs from the burgeoning studies that followed the discovery of mirror neurons. However, the two concepts are linked when we think of a possible basis for morality. We don't like to be treated unfairly ourselves and we empathise with others who are treated unfairly. We will act to correct unfairness and to prevent it recurring" (Firth 2007: 1).

wrote Martin Davies (1994), and many colleagues, philosophers and cognitive scientists agree.[9]

We first have to clear a terminological mess. Some psychologists use the term "simulation" for any kind of imaginative enacting, so it ends up as meaning: model-building and model-activating:

> The model can depict the system at some point of abstraction (...) A simulation is an applied methodology that can describe the behavior of that system using either a mathematical or a symbolic model (Sokolowski and Banks 2009: 5)

We shall use "simulation" in a narrower sense: the modelling through simulation does not primarily result in an object-depiction, but is primarily a first-person guided *process*, from which the subject can learn relevant first-person counterfactual matters (e.g. what would I do if I had to determine the price of my used car). Simulation thus involves the imagining subject (or his/her counterpart) as a part of scenario imagined. Remember: we understand others by using our own mentation in a process of simulation (Martin Davies). I shall be relying on a work already mentioned in the Introduction that is a synthesis of psychological and philosophical research on simulation (Goldman 2006). Goldman points out the existence of an alternative view of psychological understanding, namely Theory-Theory that postulates the existence of a cognitive module containing assumptions about "other minds" and ways they work.[10] He allows for combination of the two (ST stands for "Simulation Theory"):

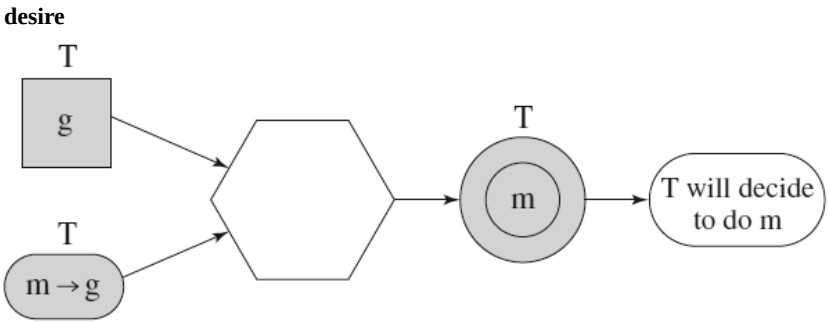
> I shall call the act of assigning a state of one's own to someone else projection. As we have just seen, projection is a standard part of the ST story of mindreading. It is the final stage of each mindreading act, a stage that involves no (further) simulation or pretense. Indeed, it typically involves an "exit" from the simulation mode that occupies the first stage of a two-stage routine. The simulation stage is followed by a projection stage. Thus, a more complete label for the so-called simulation routine might be "simulation-plus-projection". (Goldman 2006: 40)

Similarly, Perner stresses simulation but allows for Theory-Theory episodes Perner and Kühberger (2005). I would also advocate a hybrid: a blend of Simulation Theory and Theory-Theory, with emphasis on simulation. I shall also borrow a term from Goldman, "enactive imagining". He notes the following: "When I imagine feeling elated, I do not merely suppose that I am elated; rather, I enact, or try to enact, elation itself. Thus, we might call this type of imagination 'enactment imagination'" (Goldman 2006: 47). He distinguishes more primitive type of simulation, mainly unconscious and related to mirror neurons, and more sophisticated, higher kind exemplified by enactive imagining, that is

[9] See, for example Currie (2002) and the now classical text Gordon (1986). For early debate between the two kinds of theories and for important original contributions to it see *The mental simulation debate*, in Peacocke (ed.) (1994: 104).

[10] For early debate between the two kinds of theories and for important original contributions to it see *The mental simulation debate* in Peacocke (ed.) (1994: 104).

relevant to us here. The latter is characterized by its target, namely "mental states of a relatively complex nature, such as "propositional attitudes", by being partly "subject to voluntary control" and being to a high degree accessible to consciousness (Goldman 2006: 147). He proposes a nice flow chart to illustrate the simulation. Let me illustrate it with the famous Fat Man TE. I am supposed to decide whether I would push the Fat Man from the bridge in order to save five other people. Call me T I simulate my doing so; the final stage of the process looks somewhat like the following.:



Let **g** stand for "I don't want to feel guilty", **m** for "I am not pushing the Fat Man". Then (**m→g**) says that if I don't push the Fat Man, I shall not feel guilty. The decision is not to push, and the belief is my belief about myself, namely that I will not do it.

We shall use the flow chart in the sequel for most important PTEs to be discussed.

A distinction often drawn in the context of Simulation Theory is the one drawn by Robert Gordon in his (1995) and discussed by Gregory Currie in his (2002: 56ff). It concerns the contrast between two projects; in the first I "imagine is myself in your situation", in the second I imagine being you in that situation. A philosopher is immediately reminded of a puzzle famously raised by Bernard Williams in his paper "Imagination and the Self" (1976):

> It seems unproblematic for me to imagine that I am Napoleon; asked to do this, I know roughly how to comply. (Contrast this with the instruction to imagine that someone else—Abraham Lincoln, say—is Napoleon; here it is much less clear how to proceed.) But if imagining is a guide to possibility, my imagining may lead me to a further, more metaphysical thought: the thought that I could have been Napoleon. And it is this that Williams finds puzzling: "I do not understand, and could not possibly understand, what it would be for me to have been Napoleon" (Williams 1976: 45).

Indeed, how could I (or Williams or anyone other than Napoleon) have been Napoleon? Surely only Napoleon could have been Napoleon. The answer would demand a paper of its own.[11]

[11] But see Vendler (1984) and Ninan (2016) for relevant discussion.

Can we trust Simulation Theory? How certain is it that people use simulation to understand each other? Here is a cautious formulation in a recent overview, "Folk Psychology as Mental Simulation", offered by Gordon himself, together with his collaborator Luca Barlassina, in *Stanford encyclopedia:*

> In particular, while the consensus view is now that both mental simulation and theorizing play important role in mindreading, the currently available evidence falls short of establishing what their respective roles are. In other words, it is likely that we shall end up adopting a hybrid model of mindreading that combines ST and TT, but, at the present stage, it is very difficult to predict what this hybrid model will look like. Hopefully, the joint work of philosophers and cognitive scientists will help to settle the matter. (Gordon and Barlassina 2017, ST stands for Simulation-theory, and TT for Theory-Theory)

The consensus "that both mental simulation and theorizing play important role in mindreading" is enough for our purposes. We shall thus assume that the Simulation Theory is the correct theory about *a* way, possible *the most important way*, in which a human being comes to find out and understand the thoughts of her conspecifics. (This allows for other ways, like the ones proposed by Theory-Theory or special module theory.) So, we shall assume that our thought-experimenter simulates the possible states, including feelings of oneself and others, and derives her judgments from the simulation. (The presence of additional, say Theory-of-mind elements would not change the basic situation, as long as simulation does play and important role). However, I hope that most of conclusions of this paper are valid for perspective-taking and imaginative enacting in general, independently of a particular mechanism in charge of it.

2.3. *Simulation in TEs, moral, political and legal*

It is now time to bring together the issues of perspective taking and our main topic, moral, political and legal TEs. Some TEs obviously involve empathetic perspective taking that ends in sympathy with the characters involved. The Trolley and Fat man TEs are a clear example, where the experimental results show a direct and strong involvement of subjects who have to imagine, presumably enactively to push by their own hands the Fat man, and kill him in this way. We normally have no problem in simulating to some degree the pain of other. Here is what neuroscientists tell us.

> Seeing or imagining others in pain may activate both the sensory and affective components of the neural network (pain matrix) that is activated during the personal experience of pain. (Minio-Paluello, Avenanti, and Aglioti 2006: 320).

So, why people find pushing the Fat man way more problematic than just turning the switch? Apparently different parts of brain get involved. Simulation assumption might help a bit: when one simulates

turning the switch, the act itself looks neutral, apart from indirect consequences. When one simulates pushing a person from a bridge, it feels like actually doing it. The neurologists (Roth et al. 1996) tell us that in people simulating movement the primary motor cortex gets involved, as if they were themselves doing the hand movements. If this holds, there is a qualitative difference in feeling when one is simulating turning the switch, a neutral indirect causing of change in trolley's path, and when one is simulating the effortful pushing of a heavy object, the Fat man. If one feels imaginatively enacting the later as if it were one's actual effort, it is clear why it feels like killing the man with one's hands. Indeed, here simulation might explain the difference in feeling. (But more research has to be done before any definitive conclusion is taken.)[12]

The other pretty obvious kind of perspective taking are the Golden Rule cases. It is here that the very rich literature on empathy, often connected with sympathy, and sometimes distinguished, becomes relevant.[13] And simulation normally generates empathy and sympathy (see Copman and Goldie 2011).[14]

Here, the issue of moral evaluation intervenes. Let me quote the philosopher who connects morality and empathetic simulation very directly and radically. It is Mark Johnson.

> Moral imagination is our fundamental capacity to imagine how certain values and commitments are likely to play out in future experience, without actually performing those actions and having to deal with their lived consequences. The quality of our moral thinking therefore depends on (1) the depth and breadth of one's knowledge of the physical and social worlds he or she inhabits, (2) one's understanding of human motivation and cognitive/affective development, (3) one's perceptiveness of which factors are most relevant in a particular situation, and (4) one's ability to simulate the experiences and responses of other people with whom you are interacting. It is thus as much an affair of imagination as it is an appropriation of prior knowledge. (Johnson 2016: 363)

Passing to moral imagination Johnson characterizes it simply as simulation. It gives us "a deep sense of how others might experience a situation" and he connects it with empathy and talks about "empathetic imagination," which, in his view, makes it possible for us to appreciate and take up the part of others.

Let us now pass to social contract PTEs which make up one of the two most prominent kinds and traditions of macro-PTEs. (With apologies for very little space dedicated to each famous PTE in the tradition;

[12] The reader might like to consult the chapters on imagination and morals by Thomas Schramme, Antti Kauppinen, Alison E. Denham, David Shoemaker, Ishtiyaque Haji and Maurice Hamington in Maibom (2017).

[13] On Golden Rule and empathy see Neusner and Chilton (2008), Pfaff (2007) and Wattles (1996: 144ff.).

[14] Goldman has anticipated it in his (1992) again reprinted in Goldman (2013: 174–197).

I am looking for a general pattern). They are ideal for bringing simulation and political thought-experimenting together. Here, in contrast to Plato's tradition of building of the ideal state from the third person perspective, the interlocutor is asked to consider the possibility of living in some given arrangement and she is expected to imagine herself actually doing it. Here is a fine, relatively recent statement connecting the tradition to perspective taking:

> Contract theorists hold that to judge whether an action or institutional arrangement is morally justified, one must determine whether it is in conformity with principles that would be the object of agreement. They thus assume that persons are able to discern the content of this hypothetical agreement. They thereby assume, I will argue, that persons are able to determine the acceptability of principles from other perspectives than their own present point of view. This is one out of two assumptions on which my investigation regarding the empirical plausibility of contract theory will concentrate. (Timmerman 2014: 2)

Now, behind the Veil the participant does not know how rich s/he will be. S/He has to imagine him/herself being very rich (vow!), being moderately well off (not bad!) and being very poor (God forbid!). According to my general proposal, s/he uses his/her default knowledge of being rich, well off and poor. How does the knowledge then get used? Not inbuilding a further model from the third-person perspective, but in simulating: let me imagine myself being poor, etc.! Let me remind you of Rawls' formulation from his *Theory of Justice*:

> The aim is to rule out those principles that it would be rational to propose for acceptance, however little the chance of success, only if one knew certain things that are irrelevant from the standpoint of justice. For example, if a man knew that he was wealthy, he might find it rational to advance the principle that various taxes for welfare measures be counted unjust; if he knew that he was poor, he would most likely propose the contrary principle. To represent the desired restrictions one imagines a situation in which everyone is deprived of this sort of information. One excludes the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices. In this manner the veil of ignorance is arrived at in a natural way. (Rawls 1999: 16)

As we mentioned, we shall concentrate on stages two to six, stressing the third stage. At *stage one,* the question is understood by you: you realize that you have to decide on purely rational grounds, in your own best interest. At *stage two* you start consciously building the model of the scenario suggested. You might be tempted to take a risk: why not special privileges for the rich ones, at the expense of the poor ones. But then you imagine yourself being poor, and people you know suddenly being very privileged rich ones. Here the simulation might set in.

The *third stage*, we propose, concerns the production of the answer, involving the generation of intuition as to whether the arrangement is acceptable to you. This probably involves decision making at the unconscious level. Your cognitive apparatus might revive some memories of poor people that you have suppressed from your consciousness, and

they might at the end motivate you not to risk. Richer simulation helps. You then come first with an immediate, unconscious intuition (I don't want to risk, I want an I arrangement that will be generous to the poor), at the stage *four;* other consideration intervene, and at the *fifth* stage, you come out with explicit intuition at the conscious level: I don't want to risk extreme forms of poverty, I want a decent life even if I am not rich.

Let us return to Goldman's schema. Call me T. Remember, I have a belief box (of oval shape), with the relevant belief: if I reject the privileges for the rich (**m**), I might end up having a decent life (**g**), even if I am relatively low on the social scale. I also have a desire box, square shaped in the drawing. The desire to have a decent life is sitting there. My simulating apparatus, of hexagonal form, puts together the two contents, **m→g** and **g**. But how can I, the imagined or simulated T, get to **g**? Well, by **m**—rejecting the privileges for the rich.

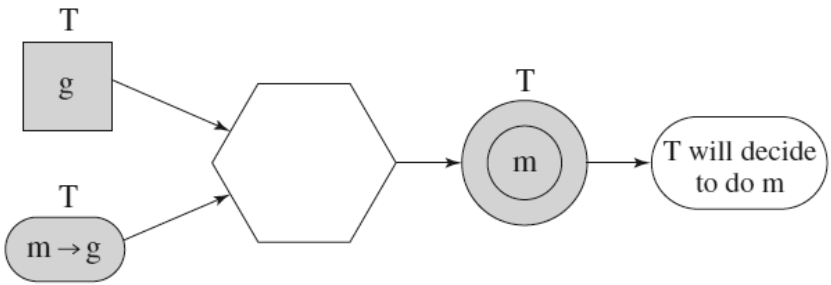


Figure 2.3. Decision attribution reached by simulation. (Adapted from Gallese and Goldman, 1998, with permission from Elsevier.)

So, T will decide to reject privileges for the rich. Rawls is vindicated.

But wait, I have assumed I shall be a well-surviving gentleman? But what if I turn out to be a female? And turn out to love children? I want more chances for myself and for them. This *sixth* stage of varying and generalizing, the intuitive induction, might make you even more egalitarian: I want children of relatively poor couple to have equal opportunities as children of relatively rich ones

What does cognitive study of simulation tell us about these processes and capacities involved? As we noted, there are two different ways of perceiving the other's situation, and these are often confused. First, you can imagine how another person sees his or her situation and feels as a result (an *imagine-other perspective*). Second, you can imagine how you would see the situation were you in the other person's position and how you would feel as a result (an *imagine-self perspective*).

Goldman notes that "egocentric" mindreading tendencies are found in both children and adults. Goldie described the imagine other perspective as "imagining the enactment of a narrative from that other person's point of view" (1999: 397). The result is not simply understand-

ing, but *sensitive* understanding. It is this form of perspective taking that has been claimed to evoke other-oriented empathic concern (Batson 1987, 1991). This imagine-self perspective connects self-recognition to other recognition (see Pfaff 2007: 65ff). A developed account of this kind appeals specifically to mental simulation or something sufficiently like it. (see Pfaff 2007: 69ff.). C. Daniel Batson, a famous author in cognitive study of perspective taking writes:

> An imagine-self perspective involves, in Adam Smith's colorful phrase, "changing places in fancy." It has also been called "mental simulation" (Goldman 1992; Gordon 1992). Especially when the other's situation is unfamiliar or unclear, imagining how you would feel in that situation may provide a useful, possibly essential, basis for sensitive understanding of the other's plight. It may provide a stepping-stone to imagining how the other is affected by his or her situation and so to empathic concern. But if the other differs from you, then although focusing on how you would think and feel in the other's situation may provide comparative context, it also may prove misleading (…). (Batson 2009: 268).

Back to Rawls: the easier task is for myself, a male with long life experience, to imagine myself being poor. It is the case of imagining myself, as I am in a different situation. The difficult task is to imaging myself being a relatively young women with with a strong attachment to my newborn child, who needs me 24 hours a day. Human beings can in principle do both Goldman's sketch of simulation offers an elegant way to depict the process (he mentions the connection (2006: 294), unfortunately without developing it).

In the situation we are discussing, I am cognitively to "quarantine" my beliefs and desires that are irrelevant (2006: 30). Let me apply it to the reasoning under the Veil. Change the meaning of **g**, **h** and the rest. Suppose that **g** stands for "I want good circumstances for my child to develop and live in", and suppose that I am a relatively indifferent male. For me, then, **~g** holds. I might also have a belief **h** that my ability to struggle and achieve good conditions for myself are way more important than social care for children. Then I will never arrive at doing **m**, say accepting a very high degree of egalitarianism.

Well, what I should do is to quarantine **~g, h, and my reservations about the m-g connection, ~(m→g) belief.** Rejecting **~g** makes me want good circumstances for my imagined child to develop and live in, rejecting **h**, helps me to avoid unreasonable self-confidence. I realize that accepting a very high degree of egalitarianism (our **m**) would provide the right circumstances for my would-be child (**m→g**):

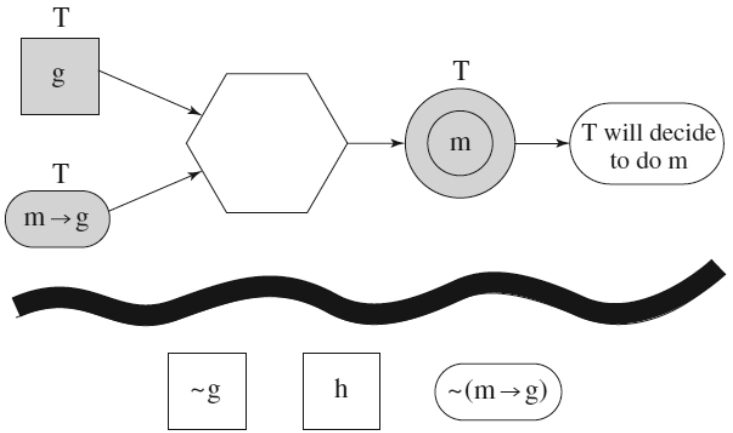


Figure 2.4. Decision attribution reached by simulation, showing quarantine.

I opt for **m**, and end up with the Rawlsian choice. All in all, I have to abstract from ('quarantine') my other interests and perhaps some relevant (male-centered) beliefs. This brings us to a frequent objection to Rawls's Original position:

> … the parties are deprived of so much information that they are incapable of making any choice at all. How can we make any rational choice without knowledge of our fundamental values? To begin with, the parties do know of their need for the primary goods and their higher-order interests in the moral powers. (Freeman 2007: 160).

Translated into cognitive terminology, how much information, in particular information about myself, can I quarantine, and still competently decide about the right choice for *me*? Let me try a hunch of an answer to the question quoted. Supposed I am behind some science-fictional veil of ignorance, let us say abducted by aliens from an inhabitable exoplanet, say Kepler 62 f, and I know two things. First, there are different societies co-existing there, some more tolerant, some less, and second, the aliens might tamper with my brain, and change the values I shall wake up with after the tempering. What society should I choose? It seems obvious to me that I should opt for the most tolerant one. In the worst case, if I am to wake up with a lot of crazy 'values' ruling in my brain, I want now to be tolerated once this happens; the more tolerant the society, the better for me. The Minimax gives the right answer: choose the society which will offer most even in the worst case. A lot more should be said, but I believe that the quarantining interpretation offers a good first step in direction of an answer.

Let me come to the end of my short list of illustration of famous TEs that seem to demand simulation in their implementation, with a brief, all too brief pointing to the work of Thomas Scanlon. One of his many examples is the right to privacy (1998: 204), but he does not give any

detailed recipe; so let me try to provide one. How would one argue for the right, discussing matters with a somewhat voyeuristic neighbor? The first move is like the Golden Rule one: one can ask the neighbor to imagine that he is being peeped upon. Imagination will involve simulation. If the neighbor sees the point, one can try to offer a more general proposal. Imagine other people, how would they feel if deprived of right of privacy? More simulation might be required. Here is Scanlon's general statement:

> Some of the most common forms of moral bias involve failing to think of various points of view which we have not occupied, underestimating the reasons associated with them, and overestimating the costs to us of accepting principles that recognize the force of those reasons. (Scanlon 1998: 206)

The simplest way to recognize the reasons associated with "points of view which we have not occupied" is by trying to simulate them. We "quarantine" our own point of view, and replace it with the target one, and then enter simulation n. (Habermas discusses "taking the attitude on the other" commenting G. H. Mead in the fundamentally important chapter on Mead and Durkheim of his *The Theory of Communicative Action* v. 2, from 1981. He then incorporates it into his own theory as its basic assumption; for a brief, principled statement see his (1995: 117)).

Let me note that Scanlon's most famous work on the topic has been done in the area of ethics; however, like other contractualists, he connects it with political philosophy and talks about morality of institutions (2016) along the same lines he proposed for individual morality.

So, back to the stages of TE, armed with a sketch of Simulation Theory. We have located the perspective taking at the stage four, the one in which the scenario proposed is being worked out. At the next, *fifth stage*, thinker comes out with explicit intuition at the conscious level, usually geared to the particular example and having little generality.

In our example, the male thinker has imagined being a female, and has arrived to the decision that the best course for him would be to opt for gender equality in the future society.

Sixth stage: since the typical job in previous stage is consideration of some particular scenario, the thinker will next have to do some varying and generalizing (deploying both moral and rational competences) at the conscious and reflective level and, perhaps, at the unconscious one too. Sometimes this process is called intuitive induction (Chisholm 1966). In our example, the thinker imagines himself as being poor, and then as being not very talented for well-payed jobs, and so on.

In the Veil-of-ignorance kind of TEs the experimenting yields a series of prudential answers-intuitions. What about the moral judgment? It is the result of more theoretical reflection, after the descriptive information gained by simulation has been systematized. (Rawls sometimes talks about a wider framework of entering social contract, with "strains of commitment" securing the moral side, but we cannot enter it now; for a fine analysis see Waldron's "Strains of Commitment" in Hinton (2015)

This kind of combined strategy is a must for the classical contractualist tradition. In Kant (and Parfit) you decide about the moral status of a maxim after you have calculated the consequences of its becoming the universal rule. In Scanlon, you decide about the moral status of your proposal after you have gone through imagining other people's reactions to it, and your attempts at persuading them. In Rawls, you decide about the normative status of your proposal after you have tested it under the Veil, possibly comparing it to other alternatives, and calculating which of them will assure the maximin result. Call the first task the descriptive-factual exercise, and the second the normative derivation. Note that Habermas and Scanlon build more normativity into the decision phase. Habermas, for example, derives it from the regulative use of speech: "The social reality that we address in our regulative speech acts has by its very nature an *intrinsic* link to normative validity claims" (Habermas 1990: 61).

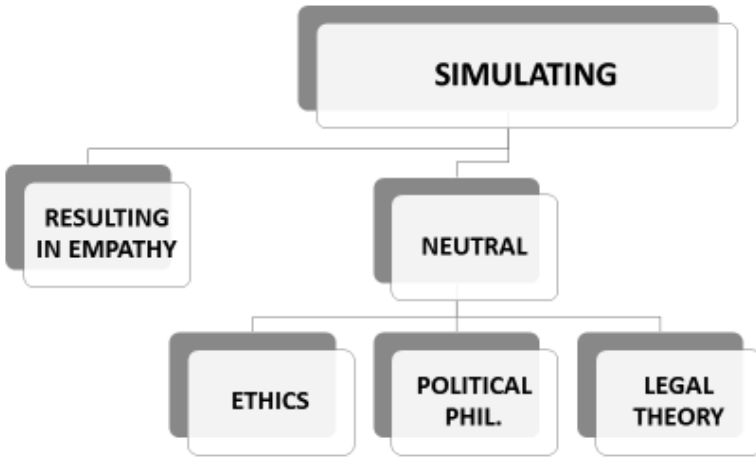
In simpler kinds of practical TEs, like the Fat Man and Golden Rule, the initial moral judgment is the direct result of empathy generated by simulation. It offers and account of how morality enters the picture, congenial to sentimentalists ethics.[15]

Let me conclude by mentioning an example from philosophy of law, the area that has not been much discussed until now in terms of thought-experimenting (but see in this issue the paper of Miomir Matulović, to whom I also owe the example that follows). Friedrich Carl von Savigny talks about the interpreter reconstructing the thought (Gedanke) "enclosed in a law". Good interpreters should put themselves in the same starting point of the legislator, and "artificially repeat in themselves his way of proceeding, so that the law may come to be born again in their mind" (1867: 171). Here, we are explicitly asked to step into ancient legislator's shoes, and pretend we are legislating in his place. This putting oneself in the place of the author ("Gleichsetzung mit dem Verfasser", in the original German) seems to be a common strategy suggested in early nineteenth century German hermeneutics. Schleiermacher asks: "But how can we understand the inner process of the writer from this? By observation. But this is based on self-observation" (1998: 135; the original manuscript dates from 1828). And, he claims that "/.../ one must put oneself in the place of the author on the objective and the subjective side" (1998: 24); the "objective" here is "linguistic" and the "subjective" is the psychological.[16]

Here is then the overview of the areas where simulation plays a central role:

[15] See for instance Slote (2007). The darker side of this connection, recently intensely discussed in connection with the Fat Man TE, is the possibility that empathetic gut reactions usurp the place of rational consideration. See for discussion and references Cushman, Young, and Greene (2010).

[16] For parallels with Savigny's contemporaries writing about understanding and empathy in general see the historical overview in the Introduction to Coplan and Goldie (2011). See also Girard (2017) and Leyh (1992).



2.4. *Should we be pessimistic about simulation? A quick glance at the literature*

What about difficulties connected with simulation? Let me briefly mention a few problems raised by some authors, then an optimistic counterproposal, and conclude with moderate optimism. The simple schemas we reproduced seem to suggest that quarantining and putting oneself in another's shoes is an easy matter, but of course, neither Goldman nor cognitive psychologists think it is. Here is a characterization of some difficulties.

Epley and Caruso (2009: 297ff) list three "critical barriers" as they call it: activating the ability to simulate, adjusting "an egocentric default" (in their parlance), and accessing information about others. For this, they have to do several things. First, they must actively think about another person's mental state, thus activating the process of perspective taking. Second, they must abstract from their own characteristics, which is normally not easy. Third, they must deploy non-egocentric information about other people in a skilled manner. (Ibid.). None of this is particularly easy.

Some authors go much further in pointing to problems. Let me choose a recent warning due to Shannon Spaulding. In her paper on "Simulation Theory" (2016a), and even more in "Imagination Through Knowledge" (2016b) she comes up with interesting challenges brings further challenges. But before doing this, she offers clarificatory and classificatory information that is extremely useful, given that the term "simulation" is used in many senses, and there is clear need to distinguish them to avoid very bad confusions; let me summarize the information quickly. Spaulding starts from Goldman's proposal according to which a process P simulates process P' if and only if first P duplicates, replicates, or resembles P' in some significant respect and two,

in its duplication of P', P fulfills one of its purposes or functions. In the case of mindreading simulation, the purpose or function of is to understand target's mental states. She then introduces the crucial distinction between abstract and concrete simulation, the first including activities like computer simulation and the like, and the second being the psychological simulation that involves "sameness of system and fine-grained process" (2016a: 264). The distinction is very helpful, and could save writers from confusions that mark the scene of present-day investigation of simulation.

Spaulding next distinguishes high-level from low-level simulation. She lists three characteristics of the former. First, it "involves imagination in the conventional sense" (267) Second, it explains our engagement with fiction, where we put ourselves "in the fictional character's position and imagine what we would think, feel, and do in that situation. Third, it explains how one can get knowledge through simulational imagination, so that if could be "co-opted to explain how some thought experiments work" (267). In contrast, in low level simulation, "imagination operates unconsciously and automatically."

Now, on the skeptical side, Spaulding's most direct challenge to the project of finding constraints that would rehabilitate imagination is to be found her, "Imagination Through Knowledge" (2016b). On her view, the puzzle of how we arrive to knowledge through imagination suggests that imagination is "not sufficient for new knowledge" (2016b: 222). The argument seems to be the following: if imagination is to be constrained by extra-imaginative pieces of information and by other abilities, then imagination does not bring new knowledge. But this is too severe a demand. Compare physical constraints. I commute from my home town to my working place about hundred miles distance. For the car to bring me to my work there should be a well-established and well-kept road, constraining the travel, there should be red lights helping to prevent crashes, and so on. Imagine someone arguing that therefore "car is not sufficient" for commuting, and is not doing any real work! Well, the fact that an item needs constraints to function properly does not entail that it never performs any function.

Spaulding has an auxiliary argument: "I have argued that the cognitive capacity to imagine scenarios is distinct from the cognitive capacities that underlie our ability to judge the accuracy of our imaginings" (2016b: 222) and "/…/there is nothing in the capacity of imagination itself that could evaluate the accuracy of the possibilities we imagine." (2016b: 222). Indeed, there is nothing in the car itself that recognizes red/green light. This does not show that the car will not take me from home to work, only that car *alone* will not do the work. So much about Spaulding's direct challenge to the instructive use of imagination.

Let me mention, however, that in her text the challenge is preceded by a rich and very provocative analysis of one particular kind of imaginational enactment, namely simulation. Her argument resembles the general one we just summarized. Her example is the following: I watch

John tease Mary, and try to figure out why he is doing this. I simulate his activity, and end up concluding that John likes Mary and is trying to get her attention. Fine, but how do I choose this option rather than some other, equally plausible in itself, for instance that he is just humiliating her? I need additional information, and my simulation tells me nothing about these matters. Again, to me it looks like simulation has done the main job, like the car in our example; the fact that the main job cannot be fully accomplished by the main agency in question, tells little against it.

So much about criticism;[17] we had to be very brief. For balance, let me conclude this section by mentioning a very helpful and more optimistic book, Peter Timmerman's 2014 *Moral Contract Theory and Social Cognition* who comes very close to our topic with an important difference—his is moral contractualism rather than the political one (and he says nothing about the psychological mechanism that makes accessible to people „other perspectives than their own present point of view.") But he has a lot to say in defense of the view that simulating oneself in various situations and simulating others are in principle within one's power.

He notes several differences between the kind of perspective-taking that normally interest psychologists, and the kind relevant for contract theorist. For example, there is the difference in the target (Timmerman 2014: 36). In contrast to psychologists who are interested in factual agreement, " we need to find out not whether others would in fact agree to principles that permit it but whether *they have reason* to do so. We are thus not first and foremost interested in what they *would* think or feel about a principle. We are, however, interested in a closely related question. As we need to determine whether others have reason to agree to a principle, we are interested in *what they would reasonably think or feel* with regard to the principle. The second difference "concerns the sort of perspectives that are considered". Philosophers are interested in general, abstract viewpoints, psychologists in our ability to recognize perspectives of "particular others" (Timmerman 2014: 36).

He further distinguishes several variables relevant for moral contract, and his picture can be easily applied apply it to the political one (Timmerman 2014: 26ff.). The first variable, he writes, "concerns which agents can use the procedure adequately to form moral judgments." A second variable, concerns *the circumstances* under which agents can apply the procedure, and the third the extent of their capacities. For all three cases, he comes close to contrasting idealizations and realistic proposals. He has some fine ideas about measures that could help normal agents to face the daunting task(s). He assumes, (...) that potential interaction partners can detect whether one can be trusted to comply or not, and as such will refrain from interacting cooperatively with persons who are not disposed to comply.) Also, he argues that

[17] See Klampfer (2018) in this issue.

we may assume that "persons are able to determine the acceptability of principles from other perspectives than their own present point of view". He mentions two important means, information gathering and "the internalization of moral principles" to which the biggest part of the Chapter Three of the book is dedicated.

Of course, the discussion between PTE-defenders and PTE-skeptics is going on, but I think we have no reason to be pessimistic about the basic abilities involved in political thought-experimenting. Let me then conclude.

3. *Conclusion*

In our investigation of cognitive mechanisms of PTEs, we have tried to bring together two blossoming traditions: the study of perspective taking and methodology of thought-experiments. Both are extremely rich, but we have narrowed our topics down to PTEs on the side of experiments, and to mental simulation on the side of perspective taking.

We have discussed the kind of PTEs that has marked the central tendency in a tradition of political philosophy, active at least from Kant on, but especially since and including Rawls' *Theory of Justice*. It is the tendency to view political justice and moral value in terms of a hypothetical contract. Political and moral TEs presented within the contractualist tradition (in the widest sense) typically ask the thought-experimenter to imagine how other people would take the experimenter's moral and political proposals, how they would feel about them, and whether and how they could be persuaded to accept them. A somewhat special but perhaps most famous case is imagining what one would propose as political arrangement if one were ignorant about one's abilities and material situation in the future situation. Again, I apologize for cramming together all the famous PTE in the tradition, each of which deserving at least a long paper attempting to account for its mechanism; but this is the price of arriving at a general pattern, if all goes well.

We have concentrated upon contractualist methodology, where imagining is supposed to yield factual intuitions about whether the subject(s) in question would accept proposed arrangements, and the normative work is done by theory. However, we have noted that there is another, more direct route to normative judgements, directly from empathy provoked by simulation, explored by a number of cognitive psychologists and stressed by Goldman on the side of philosophers; it is a matter relevant for sentimentalist ethicists, but also for understanding some very popular TEs, like the Trolley and Fat man ones. Here, the appeal to simulation yields a fine by-product, a more direct route to moral judgment.

How do all these TEs work? Our moderately inflationist mental modelling proposal is that they mobilize our imaginative capacity for perspective taking, most probably perspective taking through simulation. The framework proposed is moderately optimistic; it suggests the

answers to questions that are often raised for other kinds of TEs as well. To quote James Robert Brown, one wonders how one can learn new things without new observational data? (Brown 1991: 111ff.). In the case of our PTEs, the data come from perspective taking: the information producing capacity is either the capacity to simulate or some closely related ability. His second worry, why are thought experiments superior to deduction in terms of heuristic power, obviousness and ease, can be alleviated or even discarded by appeal to the fact that mental simulation is way more accessible to subjects than abstract political reasoning from principles and facts, and its output is usually quite obvious to the subject. The third question is: where does the "experiential" element in thought experiments come from? Are there any new experiences or quasi-experiences present in thought experiment, and of what nature are they? Yes, it is a new experience, namely the experience of simulating. In the case of empathetic simulation, the qualitative, emotional character of experience is highly prominent, and in the case of less emotional simulation, it still has experiential character ("Let me imagine that I am a generally incapable person; how would I feel in a strongly competitive society, at the bottom of its pecking hierarchy?"). Finally, as Brown puts it "if the reasoning in thought experiment is broadly inductive, how can it eliminate alternatives and reach its conclusion so quickly and effortlessly, and assert it with such force?" (Brown 1991: 111ff.). Simulation normally is quick and effortless; the simulator does not go through alternatives, but is constrained in an unconscious way.

Let us conclude by placing the account within a bigger picture, returning for the moment to our starting point. We have distinguished two kinds of PTEs and two manners of imagining political arrangements. The first consists in building third-person mental models, based on our inductive knowledge, and on default assumptions about people, about practices and institutional arrangements. The second consists in perspective taking, imagining oneself (as oneself or even as someone else) and asking about condition one would accept. Golden Rule and social contract are prime examples, either in realistic or somewhat unrealistic scenarios of ignorance and/or ideal rationality and the like

We have proposed first-person mental simulation as the basic mechanism, although we did not insist on the "purity" of mechanism. (Goldman himself proposes the idea of a „hybrid theory" according to which the simulation and the reasoning on the bases of theoretical knowledge about human minds (theory-of-mind, can interact, for example 'cooperate' (ch. 2.7 of his 2006 book); this might be an interesting option, to discuss at some other occasion. And of course, simulation might make occasional appearance in the first, predominantly first-person model building; the author might ask the reader how she would feel in such and such an arrangement, something that happens all the time in *The Republic*. But, from a wider perspective, the two mechanisms, model building and simulation, and their combination(s) exhaust the range of psychological

mechanism underlying political thought-experimenting. This is the ambitious proposal to which the present paper is a tentative contribution.

References

Batson D. C. 2009. "Two Forms of Perspective Taking: Imagining How Another Feels and Imagining How You Would Feel." In Markman, K. D., Klein, W. M. P., and Suhr J. A. (eds.). *Handbook of imagination and mental simulation*. New York: Psychology Press Taylor and Francis Group.

Brown, J. R. 1991. *The Laboratory of the Mind Thought Experiments in the Natural Sciences*. London: Routledge.

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

Chisholm, R. 1966. *Theory of knowledge*. New Jersey: Prentice Hall.

Cohen G. A. 2009. *Why Not Socialism?* Princeton: Princeton University Press.

Clohesy, A. M. 2013. *Ethics, solidarity, recognition*. London: Routledge.

Coplan, A. and Goldie, P. 2011. *Empathy Philosophical and Psychological Perspectives*. Oxford: Oxford University Press.

Currie, G. and Ravenscroft, I. 1997. "Mental Simulation and Motor Imagery." *Philosophy of Science* 64 (1): 161–80.

Currie, G. 2002. "The Simulation Programme." In Currie, G. and Ravenscroft, I. (eds.). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.

Cushman, F., Young, L., and Greene, J. D. 2010. "Multi-system Moral Psychology." In Doris J. M. (ed.). *The Moral Psychology Handbook*. Oxford: Oxford University Press: 47–70.

David D. 2018. "Art and thought experiments." In Stuart, M. T, Fehige, Y., and Brown, J. R. (eds.). *The Routledge Companion to Thought Experiments*. London: Routledge: 512–525.

Davies, M. 1987. "Tacit Knowledge and Semantic Theory: Can a Five per Cent Difference Matter?" *Mind* 96 (384): 441–462.

Davies, M. and Stone, T. (eds.). 1995a. *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell Publishers.

Davies, M. and Stone, T. (eds.). 1995b. *Mental Simulation: Evaluations and Applications—Reading in Mind and Language*. Oxford: Blackwell Publishers.

Davies, M. 2001. "Mental Simulation, Tacit Theory, and the Threat of Collapse." *Philosophical Topics* 29 (1/2): 127–173.

de Waal, F. 2009. *The Age of Empathy: Nature's Lessons for a Kinder Society*. New York: Random House.

Decety, J. and Grèzes, J. 2006. "The power of simulation: Imagining one's own and other's behavior." *Brain Research* 1079: 4–14.

Epley, N. and Caruso, E. M. 2009. "Perspective Taking: Misstepping Into Others' Shoes." In Markman, K. D., Klein, W. M. P. and Suhr, J. A. (eds.). *Handbook of Imagination and Mental Mimulation,* (New York: Taylor and Francis Group).

Freeman, S. 2007. *Rawls*. London: Routledge.

Frith, C. 2007. *Making up the Mind: How the Brain Creates Our Mental World*. Oxford: Blackwell.

Gallese, V. and Goldman, A. I. 1998. "Mirror Neurons and the Simulation Theory of Mindreading." *Trends in Cognitive Sciences* 2: 493–501.

Gauthier, D. 1986. *Morals by Agreement*. Oxford: Clarendon Press.

Goldie, P. 1999. "How we think of others' emotions". *Mind and Language* 14: 394–423.

Goldman, A. I. 1992. "Empathy, mind, and morals". *Proceedings from the American Philosophical Association* 66: 17–41. Reprinted in Goldman 2013: 174–197.

Goldman, A. I. 1995. "Empathy, Mind, and Morals". In Davies, M. and Stone, T. (eds.). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell: 185–208.

Goldman, A. I. 2005. "Simulationist Models of Face-Based Emotion Recognition." *Cognition* 94: 193–213.

Goldman, A. I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Goldman, A. I. 2008. "Mirroring, Mindreading, and Simulation". In Pineda, J. (ed.)., *Mirror Neuron Systems: The Role of Mirroring Processes in Social Cognition*. New York: Humana Press: 311–330.

Goldman, A. I. 2012. "Theory of Mind." In Margolis, E., Samuels, R., and Stich, S. P. (eds.). *The Oxford Handbook of Philosophy of Cognitive Science*. Oxford: Oxford University Press: 402–424.

Goldman, A. I. 2013. *Joint Ventures Mindreading, Mirroring, and Embodied Cognition*. Oxford: Oxford University Press.

Goldman, A. I. Forthcoming. "Mindreading by Simulation: The Roles of Imagination and Mirroring". In Lombardo, M., Tager-Flusberg, H. and Baron-Cohen, S. (eds.). *Understanding Other Minds*. 3rd ed. Oxford: Oxford University Press.

Gordon, R. M. 1986. "Folk Psychology as Simulation." *Mind and Language* 1 (2): 158–171. Reprinted in Davies and Stone 1995a: 60–73.

Gordon, R. M. 1992. "The Simulation Theory: Objections and misconceptions." *Mind and Language* 7: 11–34.

Gordon, R. M. 1995. "Simulation Without Introspection or Inference from Me to You." In Davies and Stone 1995b: 53–67.

Gordon, R. M. 1996. "'Radical' Simulationism." In Carruthers and Smith 1996: 11–21.

Gordon R. M. and Barlassina, L. 2017. "Folk Psychology as Mental Simulation." *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Zalta, E. N. (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/>.

Habermas, J. 1989b. "Towards a Communication-Concept of Rational Collective Will-Formation: A Thought-Experiment." *Ratio Juris* 2: 144–154.

Habermas, J. 1990. "Discourse Ethics." In *Moral Consciousness and Communicative Action,* Cambridge: Polity Press: 43–115.

Habermas, J. 1994. "Remarks on Discourse Ethics." In *Justification and Application. Remarks on Discourse Ethics*. Translated by Ciaran Cronin. Cambridge: The MIT Press: 19–112.

Habermas, J. 1995. "Reconciliation Through the Public Use of Reason: Remarks on Rawls's Political Liberalism." *Journal of Philosophy* 92 (3): 109–131.

Hinton, T. (ed.). 2015. *The Original Position*. Cambridge: Cambridge University Press.

Hohwy, J. 2013. *The Predictive Mind*. Oxford: Oxford University Press.

Johnson, M. 2016. "Moral imagination." In Kind, A. (ed.). *The Routledge Handbook of Philosophy of Imagination*. London: Routledge.

Johnson-Laird, P. N. 1983. *Mental Models*. Cambridge: Cambridge University Press.

Kind, A. and Kung. P. (eds.). 2016. *Knowledge Through Imagination*. Oxford: Oxford University Press.

Leyh, G. 1992. *Legal Hermeneutics: History, Theory, and Practice*. Berkeley: University of California Press.

Maibom, H. M. (ed.). 2017. *The Routledge Handbook of Philosophy of Empathy*. London: Routledge.

Metzinger, T. 2003. *Being No One The Self-Model Theory of Subjectivity*. Cambridge: The MIT Press.

Minio-Paluello, I., Avenanti, A., and Aglioti, S. M. 2006. "Left hemisphere dominance in reading the sensory qualities of others' pain?" *Social Neuroscience* 1 (3–4): 320–333

Miščević, N. 1992. "Mental models and thought experiments." *International Studies in the Philosophy of Science* 6 (3): 215–226.

Miščević, N. 2006. "Intuitions: the discreet voice of competence." *Croatian Journal of Philosophy* 16: 69–96.

Miščević, N. 2012a. "Plato's *Republic* as a Political Thought Experiment." *Croatian Journal of Philosophy* 12 (2): 153–165

Miščević, N. 2012b. "The competence view of intuitions—a short sketch." *Balkan Journal of Philosophy* 4 (2): 147–160.

Miščević, N. 2013a. "Political Thought Experiments from Plato to Rawls." In Frappier, M., Meynell, L., and Brown, J. R. (eds.). *Thought Experiments in Science, Philosophy, and the Arts*. London: Routledge.

Miščević, N. 2013b. "In Search of the Reason and the Right—Rousseau's Social Contract as a Thought Experiment."*Acta Analytica* 28 (4): 509–526.

Miščević, N. 2018. "Thought experiments in political philosophy." In Stuart, M. T, Fehige, Y. and Brown, J. R. (eds.). *The Routledge Companion to Thought Experiments*. London: Routledge: 153–170.

Neusner, J. and Chilton, B. 2008. *The Golden Rule the Ethics of Reciprocity in World Religions*. London: Continuum.

Ninan, D. 2016. "Imagination and the self." In Kind, A. (ed.). *The Routledge Handbook of Philosophy of Imagination*. London: Routledge: 274–285.

Parfit, D. 2011. *On What Matters. Volume One*. Oxford: Oxford University Press.

Peacocke, C. (ed.). 1994. *Objectivity, simulation and the unity of consciousness*. London and Oxford: British Academy and Oxford University Press.

Perner, J. and Kühberger, A. 2005. "Mental Simulation Royal Road to Other Minds?" In Malle, B. F. and Hodges, S. D. (eds.). *Other Minds How Humans Bridge the Divide between Self and Others*. New York: Guilford Press: 174–189.

Pfaff, D. W. 2007. *The neuroscience of fair play*: *Why We (Usually) Follow the Golden Rule*. New York: Dana Press.

Rawls J. 1999. *A Theory of Justice*. Revised edition. Cambridge: Harvard University Press.

Rawls, J. 1986. *The Basic Liberties and Their Priority, The Tanne(r Lectures on Human Values April 10, 1981*. In McMurrin, S. M. (ed.). *Liberty, Equality, and Law: Selected Tanner Lectures on Moral Philosophy, The Basic Liberties and Their Priority*. Salt Lake City: University of Utah Press.

Roth, M. et al. 1996. "Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study." *Neuroreport* 17 (7): 1280–1284.

Savigny, F. C. von 1867. *System of the Modern Roman Law*. Madras: Higginbotham Publishers.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge: The Belknap Press of Harvard University Press.

Scanlon, T. M. 2016. "Individual Morality and the Morality of Institutions." *Filozofija i društvo* 27 (1): 1–36.

Schleiermacher, F. 1998. *Hermeneutics and Criticism*. Cambridge: Cambridge University Press.

Spaulding, S. 2016a. "Simulation Theory." In Kind, A. (ed.). *Handbook of Imagination*. London: Routledge: 262–273.

Spaulding, S. 2016b. "Imagination Through Knowledge". In Kind and Kung 2016: 208–225.

Slote, M. 2007. *The Ethics of Care and Empathy*. London: Routledge.

Sokolowski, J. A. and Banks, C. M. 2009. *Modeling and Simulation For Analyzing Global Events*. New York: John Wiley and Sons.

Timmerman, P. 2014. *Moral Contract Theory and Social Cognition*. Amsterdam: Springer.

Vendler, Z. 1984. *The Matter of Minds*. Oxford: Clarendon Press.

Wattles, J. 1996. *The Golden Rule*. Oxford: Oxford University Press.

Williams, B. 1976. "Imagination and the Self." In *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge: Cambridge University Press.

Zwaan, Z. A and Radvansky G. A. 1998. "Situation Models in Language Comprehension and Memory." *Psychological Bulletin* 123 (2): 162–185.