
A HIERARCHY OF ARMCHAIRS: GERALD GAUS ON POLITICAL THOUGHT EXPERIMENTS

NENAD MISCEVIC

University of Maribor

ABSTRACT

The paper places the work of G. Gaus into the tradition of political thought experimenting. In particular, his strategy of modeling moral decision by the heuristic device of idealized Members of the Public is presented as an iterated thought experiment, which stands in marked contrast with more traditional devices like the veil of ignorance. The consequences are drawn, and issues of utopianism and realism briefly discussed.

Keywords: Gaus, thought experiments, ethics, heuristics, veil of ignorance

1. Introduction

Our topic in this paper is thought experimenting in moral theory and political philosophy, with the focus on Gaus's masterpiece *The Order Of Public Reason*.¹ We assume a post-Rawlsian framework, and ask about Gaus's methodology, in particular about the relative roles of TEs and factual knowledge in it. One can speak of three kinds of standardly acknowledged sources of political philosophical theorizing: intuitions elicited in political TEs, principles of general ethico-political kind, and finally, the (presumed) facts together with (presumably) true descriptive explanatory theories. Plato's *Republic* is the classic example of philosophical thought relying on the first, and somewhat on the third source, and Aristotle is his classical antagonist, criticizing the reliance on dialogical thought-experimenting (under other names, of course) and stressing the importance of factual anthropological and political material and some general principles about human sociability and

¹ I wish to thank Professor Gaus for his replies at the Rijeka conference, and the organizers and participants for the opportunity to discuss thought experiments in political philosophy.

flourishing.² Some very speculative thinkers have opted for the remaining possibility of combining principles and intuitions, and relegating facts to a theoretical back-burner.³ In contemporary philosophy few authors were willing to limit themselves to a choice between sources; the mainstream theorizing usually takes into account all three kinds of them, varying the order of their importance.⁴ In his *Theory of justice* Rawls has been quite explicit about this, linking principles to intuitions, and specifying the kind of general facts relevant for theorizing about justice. The picture changes a bit in *Political liberalism*; Gaus, in his turn, proposes a further re-distribution of roles and weights among the three sources. Political thought experiments are central to the methodology of research in political philosophy. They are normally concerned with properties of imagined political arrangement, or, more abstractly, of principles guiding and structuring them. Two great works of political philosophy, Plato's *The Republic* and Rawls's *Theory of justice* are arguably long thought experiments, and elements of thought-experimenting are omnipresent both in utopian (or anti-utopian) and contractualist thinking. (I shall in the following abridge "political thought experiments" with "PTEs", using "TE" alone for "thought experiment").

Understanding PTEs could, in longer run, improve reasoning in matters of political theory. It can also help with particular PTEs in applied matters: evaluating the relevance of the Trolley problem, or the weight of the "lifeboat considerations" in ecological ethics, and so on, with a series of crucial PTEs. In contemporary ethical and political debates the participants constantly appeal to results (and intuitions) that derive from counterfactual suppositions and even more complete scenarios typical for TEs. In order to evaluate such proposals, one should have a firm grip on methodological assumptions. In other words, we are going to explore the methodology of PTEs, because without proper knowledge of them, we cannot expect to figure out the relevance of the actual TEs which are important in themselves and crucial in the moral-political debate.

In his explicit formulations Gaus is often critical about thought experiments and somewhat pessimistic about their reach. He points out that "Rawls increasingly worried that as the abstraction is undone and people come to know their comprehensive conceptions of value, their devotion to the principles might be "overridden" (§3.2). Full knowledge of evaluative standards may change what is validated from the perspectives of citizens" (336). He develops Rawls's reservations further and in new directions. His criticism seems to point to the need of facts, and more facts, and in this sense he seems to be playing Aristotle to Rawls's Plato. I want to argue here that this appearance is misleading: the new factual knowledge (about social facts, human nature, game theory, and so on) leads in his story to new rounds of thought-experimenting. While Plato and Rawls have been offering relatively simple mega-PTEs, piling up the results of micro-PTEs, each concerning some particular arrangement, Gaus offers a mega-PTE in which we are invited to climb the ladder of reflexivity: first-order PTE, than new facts, then second-order PTE, and so on. We are invited to experiment by imagining other,

² A separate question concerns the status of religious data in the political thought of Islamic and Christian authors, like Al-Farabi, Augustine and Thomas Aquinas. I am inclined to treat most of it as assumed facts, and the rest as principles roughly on the same footing with usual moral principles.

³ An example could be Hellenistic neo-Platonists and their medieval followers.

⁴ See for instance the work of G. A. Cohen, most recently his (2008) volume.

somewhat idealized thinkers experimenting in their turn; this is indeed a challenging and innovative form of thought-experimenting, not proposed in the extant literature. Interestingly, it looks like an original kind of revival of dialogical PTE, absent from political philosophical writing for centuries.

Let me then, in the remainder of the *Introduction*, summarize my view of PTEs in general, in order to provide a background for discussing Gaus's innovative proposal. Then we turn to Gaus; in section II his criticism of Rawlsian abstractionist methodology of thought-experimenting is presented and briefly analyzed. Section III is the central one of the paper. The conclusion raises the question for Professor Gaus: does he agree with our understanding of his methodology and how would he himself characterize it? Let us now pass to setting out the general framework for the debate.

What is a PTE? Roughly, it is an episode of thinking that involves appeal to imagined, counterfactual situation in order to answer a moral-political question. On the most general level, features of political TEs are the following:

- (a) Thought-experimental reasoning involves reasoning about a particular set of social and political circumstances, which may be specified in more or less detail,
- (b) The reasoner's mode of access to the scenario is via imagination rather than via observation.
- (c) Contemplation of the scenario takes place with a specific purpose: forming judgment about some politically relevant theoretical proposal. (This characterization transposes to PTEs the general proposal of Tamar Szabó Gendler (2004), who concentrates upon TEs in science)

However, more precision is needed. Does any act of imagining count? The proposal is to distinguish a wide and a narrow conception. On the wide conception, even a small imaginative experiments, like those involved in finding out a categorical imperative concerning some given action (token and type), i.e. imagining everyone performing the action, and considering the consequences count as TEs. On the narrow conception only systematic, worked-out counterfactual scenarios count as TEs. The ambiguity is not peculiar to TEs, but rather comes from the very wide application of the term "experiment". The Merriam-Webster Dictionary offers as its first example sentences that depicts scientific practice, talking about "simple laboratory experiments" and "some experiments with magnets". But then it passes to "an experiment in living more frugally" and concludes with "the city's experiment with a longer school year". These last two are not the province of philosophy of science for sure; and we have no problem to distinguish full scientific meaning of "experiment" from its less demanding relatives. For us the most relevant distinction is one between rudimentary experiments or quasi-experiments (available even to children) and full-fledged scientific ones. We might draw the parallel: rudimentary (quasi-)experiments vs. full-fledged scientific ones on the scientific side, and rudimentary (quasi-) TEs vs. full-fledged philosophical ones on the scientific side. The rough-and-ready tests, like for instance Hare's put-yourself-in-the-other-person's-shoes test are rudimentary in this sense. The golden-rule test is

the most famous rudimentary (quasi-) TE in ethics. Kant's universalizability test in its simplest form is certainly rudimentary, but in the (recent) Kantian literature it has been developed into a more complex strategy. Indeed, less sophisticated procedures that don't qualify as full-fledged TEs also yield intuitions, but this does not entail that PTEs are not a useful, we shall later argue even an indispensable tool for the proper evaluation of existing political arrangements or institutions. The right parallel is between the two pairs: rudimentary TEs vs. full-fledge ones and rudimentary testing vs. controlled testing in science. A rudimentary proto-experiment with my marker can convince me that it will not leave inerasable traces on the white board, but it is a fair cry from controlled testing in the factory, which is itself a fair cry from high-level physics experiment in CERN.

Political TEs conform to a large extent to the pattern of TEs in other domains, like theoretical philosophy and science, with one important difference: the important political TEs are often "macro-thought experiments", encompassing a lot of smaller TEs, testing the justice and other moral properties of particular political arrangements, and organizing them into a global political arrangement. (And the similar structure is often found in fictional utopias, positive and negative, in what officially counts as "literary works of fiction"). Note that the Republic is a dialogical TE, in which debate does play a role; typical contemporary TEs are not meant to be dialogical, and it is an interesting feature of Gaus's proposal that it does bring back some of the old dialogical format, or at least I shall argue that it does.

Here is my view of the typical structure of PTEs, micro- and macro (see Miscevic 2012 for more detail). Usually, at stage one, a question is asked about a particular arrangement. For example, in the Republic Socrates asks whether the community of children is just, thereby prompting his interlocutor, say Adeimantos, to consider or even concretely imagine the scenario. At stage two, the question is understood, one hopes correctly, by the interlocutor. At stage three is placed the attempt of the thinker actually to imagine the situation to be judged, her tentative conscious production, the building of the "model" of the scenario at the conscious level. The fourth stage is more demanding. It concerns the production of the answer, involving the generation of intuition as to whether the arrangement is just or unjust. This probably involves reasoning at the unconscious level. At the fifth stage, the thinker comes up with explicit intuition at the conscious level, usually geared to the particular example and having little generality). If the consideration of a particular scenario is typical, the thinker will have to do some varying and generalizing (deploying both moral and rational competence) at the conscious and reflective level and, perhaps, at the unconscious one too. Sometimes this process is called intuitive induction (e.g., "What if the children were Spartan or Mycenaean rather than our own? " The arrangement is valid for all"). This is the sixth stage.

Stage seven finally brings general belief at the reflective level ("All children in every ideally just polis should be common to all citizens"). Three components are prominent in such a procedure: first, the aggregation of micro-TEs; second, the harmonization of the results of these micro-TEs and finally, the judgment of their coherence with other moral intuitions one might have. In other words, the philosophical unification can be

described in terms of narrow reflective equilibrium. This is stage eight. One can call the harmonious unification of micro-TEs “the topical narrow reflective equilibrium” and the final narrow result “the general narrow reflective equilibrium”. The former, topical one is geared to the unity of narrative structure, plus the relevance and coherence between particular stages, of the micro-TEs.

The attention to the internal structure of TEs is important for the purpose of discussing their viability and value, since criticisms of them often result from the failure to understand the specific requirements and liabilities of each stage. This will become dramatically obvious in what follows, when we address our main topic.

2. Gaus: the Limits of Abstractive PTEs

We are ready now to consider Gaus’s impressive *Order of Public Reason* so, (in particular chapters V, VI, VII). The work is placed squarely in the Rawlsian tradition of public reason liberalism and Gaus notes the tension present in Rawls between the veil-of-ignorance TE and his later stress on public reason, although in other context he speaks of “thin and uncertain” line between

A Theory of Justice and *Political Liberalism*. (2004, 111) The TE has to declare irrelevant “much of what a person understands as basic to her evaluative outlook” (42; the pages of *The Order* shall be indicated by number in brackets). The problem is even more general than Gaus notes. Besides the issue of the possible unreliability of one’s political imagination, and the irrelevance of distant possibilities which border on the impossible, there is the problem of selection. When depicting the scenario to be imagined, one has to import various assumptions, in which one is necessarily selective. For instance, Rawls was not sufficiently explicit about the status of women and the issues of “race” and the criticisms this has provoked, which together suggest that he unintentionally imported untenable assumptions about the functioning of society into his TE. Each great philosopher imports some assumptions or forgets to import others that his critical readers find essential.

But then, what justifies the results? “Rawls increasingly worried that as the abstraction is undone and people come to know their comprehensive conceptions of value, their devotion to the principles might be “overridden” (§3.2). Full knowledge of evaluative standards may change what is validated from the perspectives of citizens” (*Ibid.*). Rawls saw this stability problem, and so insisted that the abstract justification is only pro tanto — full justification must admit the full range of relevant evaluative considerations, which might override the abstract justification.” Moreover, Rawls’s experiment involving original position is individualistic, not seriously dialogical. He tried contextualizing and rooting his PTE: veil of ignorance started as universally valid but then Rawls discovers that it is not contextualized it into the existent liberal tradition and might be seen as too ad hoc. So, his later liberalism of public reason stresses the need of grounding one’s judgments in an independently available tradition.

Gaus’s proposal is distinctly post-Rawlsian, focused upon the tension between abstract TE and the concrete liberal tradition(s). It also addresses another problem, namely the role of empirical information. Rawls himself famously provided one possible answer

in his demand for wide reflective equilibrium, meant to bring in information from HISTORY, PSYCHOLOGY and SOCIAL SCIENCE. Gaus goes much further in specifying the role this information is to play in his socially oriented political and moral theory. Consider Gaus's reconstruction of Rawls's original TE:

Rawls' argument from the original position can be understood as a formalization of this two-step bracketing procedure.

First (via the veil of ignorance), we abstract away "private ends" that would lead us to legislate. One excludes "knowledge of those contingencies which set men apart."

Second, we attribute to the parties a concern with primary goods that provide a basis for their common deliberation. (37)

Initially, Gaus offers a Rawls-inspired abstraction-centered PTE, that is meant to specify the minimal initial format of the liberal arrangement, namely the famous "Rights of the Moderns". Once this is done, we are invited to go beyond Rawlsian thought-experimenting. Let me briefly summarize the abstraction-centered phase, before passing to the second phase.

Here is what Gaus calls the "Deliberative Model": he invites us to imagine slightly idealized "members of the public" who deliberate about possible rules. We are to imagine them deliberating at their armchairs, make them a bit more concrete than, say, fully informed and perfectly rational agents, but still better than most of us. Their task is divided in two parts, the first part being quite close to the Rawlsian project. The Members of the public should "abstract away from our many disagreements to our fundamental common evaluative standards as agents" (390). In this, their views and standards will be similar to those available from the single armchair. They will come up with rights of agency as basic protections from interference, manipulation, and invasion, as well as assurances that certain assistance will be provided or goods supplied. Also, some jurisdictional rights, say those concerning private property might be found by abstracting "away from this diversity to discover some common basis for evaluation of moral rules" (335). Gaus argues that this much will be stable under full justification: once free and equal individuals are aware of their full range of relevant values and concerns, they will agree that the basic Rights of the Moderns are in their interest. But, very soon, we face the limits of arguments from abstraction. "The problem then is how to continue on to the justification of more specific moral rules. At this stage, comprehensive conceptions and the freestanding justification are apt to interact in complex ways." (42)

3. Imagining the Members of the Public

Gaus thus joins those theoreticians that argue that general principles are just too general, incapable of dictating their own application. One needs to derive norms and rules, and the question is how to do it. We saw that some rules can be derived with the help of abstraction. But how does one derive the rest? One demand is clear: we need more empirical information. But how do we integrate it with other two kinds

of sources, the intuitions and the principles? Some authors eschew the demand for at least the first of the two, some, more radical, like for instance Michel Foucault, just rest content with historical information and hints for possible conclusions. Gaus is far from this kind of minimalism. And he also stresses the need to take into account possible deep disagreements that go beyond the minimum guaranteed by abstraction. Here is what he proposes

Members of the Public are confronted by a set of socially optimal eligible interpretations of an abstract justified right, just as they were confronted by a socially optimal eligible set of possible rules, x, y, z . We have compelling reason to think that the set would be neither null nor a singleton. Their justificatory problem is narrowed (regarding a range of issues, only rules that are interpretations of the abstract right are now proposed, and they agree on the central importance of having a socially eligible set), but there is no reason to suppose it will have an agreed-upon optimal element. (370)

The Members of the Public then propose specific interpretations of these abstract rights and rank their proposals as well. Since they are good-willed and rational agents, reflecting on their entire set of concerns and we consider the rules that have been accepted by them; those that have been have also passed the test. Now, we philosophers ask, sitting in our armchair: how will these members deliberate? Remember that these “Members of the Public” are “the rationalized counterparts of real moral agents” (267). The answer is that each of these imagined Members of the Public, consults then only her own evaluative standards and sees that acting on this moral demand is the thing that best satisfies those standards. And a give norm is a rule of social morality only if each and every Member of the Public endorses it as binding . We, philosophers, put their reflections together and this of course, brings back intuitions, hopefully, intuitions informed by sufficient empirical information. We thus have a plurality of imagined armchair thinkers, brought to life by one meta-reasoner. Gaus himself in chapter V brings in the Kantian metaphor of the Kingdom of Ends (paragraph 15.1 bears the title *Modeling Legislation In The Realm Of Ends*) to describe the scenario imagined, thus revealing the fact that the degree of idealization proposed is quite dramatic.

It seems that the original PTE is now being iterated: instead of merely questioning oneself about the rule(s) to be adopted, the thinker enters a 2-steps procedure. In brief, we are dealing here with an extended PTE in the service of Public reason liberalism. This is then a proper continuation of single-armchair TE. Remember the stages of macro-PTEs listed in the Introduction. Suppose we have gone through all of eight stages of the initial abstractive PTE: understanding particular questions, imagining the scenario, answering (in several steps), generalizing and equilibrating. It seems that we now have stages nine to twelve: imagining a group of deliberators being asked a question, understanding it, imagining in their turn possible scenarios and then coming up with acceptance or rejection of the rule. Perhaps we can describe this procedure as a more dialogical one than the typical modern TEs. The dialogue is more complex than the classical Platonic ones: instead of discussing matters with his interlocutors, the thinker imagines characters who deliberate together, and then s/he, the thinker, reaches

conclusions from their imagined deliberation.

Gaus goes on to apply game theory to the imagined deliberation and interaction. He asks us to suppose that Members of the Public have refused to appeal to *Deus ex machina* philosophical device for solving their problem; instead of this they “stop deliberating and, embracing their results (the socially optimal eligible set), they begin to interact p 393. In the interaction they” face an impure coordination game” (*Ibid*); Gaus then explains how the game might proceed. The game-theoretic knowledge is thus used *within the imaginary situation*; it is being imported into the extended TE, and used there to predict the behavior of the imagined Members of the Public. Let me just remind the reader that Gaus ends up with an interesting and optimistic conclusion:

There can be a procedure O, which uniquely selects x from the optimal eligible set, and although O is not itself publicly justified, each Member of the Public, consulting only her own evaluative standards, has sufficient reason to act on x. If this is so, then x solves the fundamental problem; each adopts rule x as the moral rule to be followed, and each has sufficient reasons to do so. Thus x’s authority is consistent with the freedom and equality of all. (392)

In consequence, the type of moral equilibrium sketched is one that respects the moral freedom and equality of all. “ Each Member of the Public, when confronted by a moral demand consults only her own evaluative standards and sees that acting on this moral demand is the thing that best satisfies those standards. ” (401). The whole procedure sounds rather armchair-like, a proper continuation of our eight-stage schema. We should thus distinguish levels: on the first level, the members of the public have empirical information about the particular path-dependent solutions available. On the second level, the theorist in her armchair, has the general knowledge that the members of the public will have some such solutions available, and she offers a general, meta-level scheme of justification for the decisions they might make. The whole process is a second level TE.

However, this is not the end of the story. Here is a possible objection to the present reading of *The Order of Public Reason*. Gaus is keen on actual history, and insists that contingent history can be justificatory; his example is the signing of Magna Carta, enabled by the vicissitudes of crusades, but bringing far reaching positive consequences in its wake. He notes that “a social equilibrium on a member of the optimal eligible set provides the basis for all to conform to a social morality while she still obeys only herself” (418). He goes on to note that this equilibrium is justificatory. However, “(H)ow we have arrived at this rule is a combination of contingent history, moral ideas, happenstance, and the exercise of power. The route to it is path dependent” (*Ibid*.). So, the objection goes, it is factual history that might decide, not thought-experimenting.

Answer: we should not forget that the enterprise is a second level TE. The path-dependence is engraved in the mentality of the Members of the Public: since their ancestors have signed, say, Magna Carta, they trust Parliament more than they trust the Monarch. Therefore, the theoretician has to count with this preference when she herself

is imagining their deliberation. Or, to take a more challenging example: Gaus is sure that difference principle is a quite utopian rule (444); but this may be due to the fact that the last two centuries of American history have been marked by very fast growth of social inequalities, that made a prioritarian society unthinkable. The Members of the Public in a Scandinavian country might have had quite different experiences, and this difference has to be taken into account when proposing a set of rule to be enacted in Iceland.

The wider epistemological background of this issue has to do with armchair scrutinizing of possibilities (“possible worlds”). Once the actual historical path has been described (say, the history of self-management in my country, Croatia, then part of Yugoslavia), the moral theoretician should count with preferences and mentalities co-determined by this path, and can then, in a first- or second-level armchair scrutinize the moral consequences regarding the acceptable systems of rules for my co-nationals. Scrutability from factual basis is not unproblematic. Philosophers like Jackson (1997) and Chalmers (2012) go all the way: once you can survey (“scrutinize”) the worlds, you also get the information about moral value, since all truths are scrutable from base truths.⁵

We can then say that the decisions of the Members of the Public are empirically grounded, but the second level proposals of the theorist are armchair ones, guided by general theory of rationality (game theory and the like), and including limited information about actual paths taken by history. So, let me rephrase Gaus’s proposal as I see it: what can be done in the simple armchair is to ground the right of the moderns after that level, we go more abstract. The liberal theoretician justifies an abstract schema of decision making, and he does it considering from his armchair the members of the public making their decisions. The abstract schema is sufficient for a general liberal theory; the rest is in the hands of contingent paths taken by history and offering themselves to the Members of the Public. We do stay in PTE, and empirically informed, but still situated in philosopher’s armchair.

Before passing to the conclusion, let me note a problem for the model, returning to the example of difference principle. In Chapter 7, § 21.4 entitled “The Dangers of Utopianism” Gaus warns the reader from making too idealistic demands on public reason.

The utopian aspiration also itself leads to injustice. We deeply disagree

⁵ The idea of scrutinizing possible scenarios is developed in great detail in Chalmers (2012) on almost 500 pages, and the issue of getting to the moral matters from physical-plus-mental matters is raised. Chalmers asks: are moral truths scrutable from non-moral truths? And if they are, he claims, then moral truths will not pose a distinctive problem for scrutability. Here is the gist of his answer.

On the face of it, there are good grounds to hold that insofar as there are moral truths and they are knowable, then they are scrutable from non-moral truths. Certainly, given that moral truths are knowable at all, they appear to be empirically scrutable and conditionally scrutable: given full enough knowledge of the non-moral properties of a situation, we are in a position to know its moral properties. ... Moral truths will be inscrutable from non-moral truths only if some crucial principles or conditionals governing inferences from non-moral truths to moral truths are unknowable, or if any inference from non-moral truths to moral truths has an irreducibly empirical justification. But there is little reason to believe in unknowable moral principles here, and there is little reason to believe in such an irreducibly empirical justification. (2012, 264)

about ideal justice. Rawls and his followers have painted a portrait of the fully just society, but it is not the picture of all free and equal persons. For Rawlsians to insist, in the face of the intense and sophisticated disputes about the nature of distributive justice, that all free and equal persons endorse this specific vision as their own ideal simply cannot be accepted as credible. Free and equal persons disagree about ideal moral arrangements. (445)

The first sentence sounds reasonable and familiar from a long tradition, figuring authors like Popper and Berlin. But the reader might be surprised that difference principle would be such a dangerous utopian principle potentially leading to serious injustices. The fact that many rich and moderately rich people would not agree to limiting the growth of their richness to the fate of the worse off members of their society might be taken in two ways: either, as Gaus takes it, as an objection to difference principle, or, as a more leftist theoretician might see it, as making these people ineligible as they are, to the role of a Member of the Public. Indeed, the Members are a bit idealized: why not take idealization to involve curtailing some of the natural egoism?

Let me generalize a bit. Imagine interviewing intelligent, well educated and well-meaning ancient Athenians about the institution of slavery. Judging from Aristotle and the dramatic writers, most of your interlocutors would stick to treating it as just, and finding some way or other to justify it. Most of them would propose that slaves should be treated better than they were actually treated, perhaps much better. But abolition would certainly be out. Conclusion: society without slavery would be a dangerous utopia, beyond the reach of Athenian Members of the Public. Doesn't sound just. Gaus could retort that slavery is already out at the stage considered thanks to the abstraction strategy and the list of basic human rights constructed with its help at the previous stage. But if we can proceed so decisively and brusquely when it comes to slavery, why not in the case of dramatic inequality? Alternatively, Gaus could propose a more general schema of distributional equality, that might get the votes of the preferred committee of Members of the Public, and this would be a proposal that I would gladly welcome. Otherwise we would face a serious problem: any systematic arrangement that seems just to us philosophers, but cannot be expected to be accepted by many of our less theoretically interested compatriots would appear as dangerously utopian. And every concrete arrangement that does not have full blessing of all our colleagues would be equally dangerously utopian. Question to Professor Gaus: *what should be done?*

Let me now return to the big picture.

4. Conclusion and Final Questions

There are two major and conflicting methodological options concerning political TEs. The first, in favor of the armchair, claims that they are essential given the complexity of social life, and that imagining relatively distant possibilities teaches us about what is morally required, by separating in imagination the morally relevant from irrelevant aspects that cannot be so clearly separated in empirical research bogged down in actual

contingent details of historically given political arrangements.⁶ The opposite view points to serious problems with TEs, and suggest that we don't use them, or at least that we don't put much justificatory weight on them. Some passages of *Order of Public Reason* are quite critical of the Rawlsian thought-experimenting. However, we have argued that Gaus in fact proposes a novel scheme for PTEs, namely building a two stage model in which the theoretician reflects about the imagined reflection of her Members of the Public. Empirical information (ranging from history to more speculative matters like decision theory and game-theory) is integrated into the assumptions about the Members of the Public, but the ultimate verdict is in the domain of the armchair. If successful this methodology could mediate between the more abstract thought-experimental proposals and the more tradition-based ideas of public reason liberalism, as well as between the armchair and the empirical scientific evidence.

My main question for professor Gaus is therefore the following: *do you agree with this diagnosis about your methodology, and do you see the methodology as the second-level armchair reflection?* In his immediate reaction at the conference, Professor Gaus's kindly answered to this question pointing to different perspectives governing political reflection, quite along the lines of the conclusion of Chapter V of his *Order of Public Reason..* On the one hand there is "you and me and our morality"; being philosophers we construct a model of an agent, I presume idealized, and then go on experimenting with the model. On the other hand, there is the need to consider the perspectives of the actual people in society, and for this we have to look at actual society. Of course, this is not the end of the story, since in the book Gaus does integrate the information about actual society into the imagined model prominently figuring the Members of the Public. So, let me ask the question now in a written form, hoping for a positive answer.

I see Gaus's reiteration of thought-experimenting on 2nd level as pointing to the indispensability of PTEs: a TE is an original and irreplaceable "representational device", which does not reduce to mere illustration of some piece of abstract knowledge or theory. My own preferences are of course in favor of political TEs. In science the armchair plays a crucial role (controlled by other devices, from microscopes to telescopes), so why not in political philosophy?. Given their ubiquity in the history and present scene moral philosophy probably could not exist without TEs. The critic may respond that, it doesn't follow from the fact that some tool has been used frequently for some task in the past, that it is the most efficient tool, let alone indispensable, for that job. Even granted the proposed claim that there is only one real alternative to PTEs (i.e. arguing

6 How can imagination, deployed in successful PTEs open the door to the realm of normative truths? Or even imagination plus desire? The view to be defended is that in good case the intuition does track some objective moral quality of the target arrangement or action; the preferred view about objectivity is that it involves appeal to how idealized observer would react to the target item. Take the goodness-as-experienced, goodness. Such goodness is being intentionally experienced as being a property of an action, or state of affairs. (A transparency datum). I would briefly argue as follows:

The experiencing, in the case of imagined scenarios, is an act of moral intuition.
But experienced goodness of a political arrangement is not an experienter-independent property of action, or state of affairs. (From science)

Goodness is not a property of subjective state (From Transparency).

Therefore (by principles of charity and by inference to the best explanation),

Being good in objective sense is being such as to cause the response of experiencing goodness in normal observers under normal circumstances.

Since goodness and value in general is thus response-dependent, we can come to know about it from the armchair. (see also Gaus, (1990), in particular the section "Color analogy").

from general principles and the relevant empirical data), the PTE tradition would have to be able to show better results than its rival tradition. But what are the relevant criteria here and would the comparison come out in favor of the PTE-tradition? But what about the worries? Well, the controlling devices are available here as well, and they enter the wide reflective equilibrium; Gaus's two-stage proposal might offer us some more, and thus further confirm the reliability of the well-balanced politics from the armchair.

REFERENCES

- Chalmers, D.J. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Cohen G.A. 2008. *Rescuing Justice and Equality*. Harvard University Press.
- Gaus, G. 1990. *Value and Justification*. Cambridge: Cambridge University Press.
- Gaus, G. 2004. "The Diversity of Comprehensive Liberalisms" in *Handbook Of Political Theory*, eds. Gaus, G. and Kukathas, Ch, Sage.
- Gaus, Gerald. 2011. *The Order of Public Reason, A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.
- Jackson, F. 1997. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.
- Miscevic, N. 2012. Political Thought Experiments from Plato to Rawls. In *Thought Experiments in Philosophy, Science, and the Arts*, ed. Frappier, M, Meynell, L. and Brown, J. R. London: Routledge.
- Miscevic, N. 2012b. Plato's *Republic* as a Political Thought Experiment. *Croatian Journal of Philosophy* 12 (2): 153-165.
- Rawls, J. 1971. *A Theory of Justice*. Harvard University Press.
- Szabó Gendler T. 2004. Thought Experiments Rethought—and Repercived. *Philosophy of Science* 71 (5): 1152-1163.

Received: June 9, 2013

Accepted: June 24, 2013

Faculty of Philosophy
University of Maribor
Koroska cesta 160, 2000 Maribor
Slovenia
vismiscevic@ceu.hu