Formal Models in Normative Political Theory

Hun Chung[1]

Brian Kogelmann[2]

1. Introduction

Most political scientists are now well-acquainted with the distinction between "positive" and "normative" political theory. Positive political theory, which emerged in the late 1960s from the Rochester school of political science led by William H. Riker, sought to revolutionize the field by "placing political science on a scientific footing" (Amadae and Bueno de Mesquita 1999: 273).[3] Superficially, the terminology suggests that the divide centers on each discipline's fundamental objective: normative political theory aims to *prescribe* how political affairs *should be*, whereas positive political theory seeks to *explain* or *predict* how political affairs *actually are* or *will be*. However, proponents of positive political theory did not distinguish their field from normative political theory solely based on theoretical objectives; they also differentiated it in terms of methodology. Consider the following:

[1] Department of Quantitative Theory and Methods, Emory University, Atlanta, GA, USA, and Faculty of Political Science and Economics, Waseda University, Tokyo, Japan.

[2] Department of Philosophy and Department of Political Science, Purdue University, West Lafayette, IN, USA.

Corresponding Author
Brian Kogelmann, Department of Philosophy and Department of Political Science, Purdue University, 100 N. University St., West Lafayette, IN, 47907. Email: bkogel89@gmail.com.

[3] In this way, the origins of positive political theory were similar to the early logical positivists' aspiration to demarcate "science" and "pseudo-science."

> Positive political theory, narrowly understood, means rational choice theory applied to the study of politics … Its limits are set by two familiar contrasts: positive, or what is, is contrasted with normative, or what ought to be (Forbes 2004: 57).

> Positive political theory studies politics with mathematical methods. Positive political theory is explanatory or predictive, rather than normative or prescriptive (Diermeier 2015: 1).

> I describe the field in which I expect to be working ... as 'formal, positive, political theory' ... By formal, I mean the expression of the theory in algebraic rather than verbal symbols. By positive, I mean the expression of descriptive rather than normative propositions (William H. Riker as quoted in Amadae and Bueno de Mesquita 1999: 276).

We can see here that the use of formal and mathematical methods was definitionally linked to positive political theory, which was supposed to be in sharp contrast with the informal, non-mathematical (and often deemed "unscientific") methods employed in normative political theory. The propagation of positive political theory so defined had a significant impact on spreading the rational choice approach to politics, and eventually became mainstream in the field of political science. But, at the same time, dividing positive political theory and normative political theory definitionally in terms of their methodological approaches had a rather unfortunate consequence of making many political scientists hastily conclude that formal and mathematical methods are suitable *only* for positive and *not* for normative political inquiry.

We challenge this widely held belief in the current paper. There *is* a role for formal tools and methods in normative political theory. In academic political science as well as philosophy journals it is now not uncommon to find formal models of Thomas Hobbes's and John Locke's

states of nature, of John Rawls's and Robert Nozick's theories of distributive justice, of normative theories of democratic deliberation, and more.[4] When they don't use formal tools, normative political theorists often reason with *informal* models that can and—we shall argue—should be formalized (Johnson 2014; Knight and Johnson 2015: 38-41).

Why should political theorists use formal models in their work? Formal models can be used for many purposes such as explaining, isolating, predicting, organizing, theory building, enhancing conjectures, exploring concepts, and more (e.g., Clarke and Primo 2012: 83-93; Mershon and Shvetsova 2019: ch. 2; Page 2021: 15). Speaking in the broadest and most general terms possible, however, building a formal model is valuable because doing so "forces precision in terms of one's argument" and thereby "contributes to clarity" (Fiorina 1975: 136). There are at least three ways normative political theory can benefit from this increased precision and clarity, we shall argue.

First, formal models can make rigorous an important tool in political theorists' toolkit: thought experiments (§2). Second, political theorists often propose complex theories of justice and/or democracy that typically consist of multiple normative principles; formal tools can help determine the consistency of these principles and hence the tenability of these theories (§3). And finally, political theorists often design novel institutional arrangements in hopes these arrangements satisfy certain normative desiderata. Because there is no empirical record to

[4] We discuss much of this work below. For an incomplete list of papers and books we do not discuss see Roemer (1996); Roemer (2004); Tungodden and Vallentyne (2005); Moreno-Ternero and Roemer (2008); Vanderschraaf (2010); Braham and van Hees (2014); Patty and Penn (2014); Bruner (2015); Ingham (2019); Ingham and Lovett (2019); Kogelmann (2019); O'Connor (2019); Vanderschraaf (2019); Barrett (2020); Bruner (2020); Chung (2019); Chung (2020); Chung and Duggan (2020); Chung and Kogelmann (2020); Motchoulski (2021); Schaefer (2021); Chung (2022a; 2022b); Ingham and Lovett (2022); Juarez-Garcia and Schaefer (2022a).

measure novel institutions against, formal models are a good option for gaining insight on how these institutions might work in practice (§4).

Though normative political theory can benefit from formal models, incorporating these tools into the field is not without difficulties. The first challenge concerns evaluating formal normative models (§5). What makes one good or bad? The second challenge concerns the effect an increased use of formal techniques may have on where political theorists allocate their research efforts (§6). To conclude, we consider how embracing formal theory might change for the better political theory's status in the discipline of political science (§7).

2. Thought Experiments

Although thought experiments are used in all subfields of political science, they play a central role in normative political theory. Indeed, it would not be too incorrect to say that the use of thought experiments is the political theorist's main methodological tool. What, exactly, is a thought experiment? There are many answers available, but in what follows we shall rely on the definition developed by Kimberly Brownlee and Zofia Stemplowska. They write:

> A *thought experiment* is a multi-step process that involves (1) the mental visualization of some specific scenario for the purpose of (2) answering a further, more general, and at least partly mental-state-independent question about reality (Brownlee and Stemplowska 2017: 25).

For example, to figure out what principles of justice should govern our basic political and economic institutions, Rawls asks us to imagine what kind of society rational and reasonable persons would choose to live in if they didn't know their race, gender, social class, generation,

religion, natural talents, and so on (step 1). This is known as the *original position*. What individuals choose in the original position answers the question of what principles of justice are appropriate to govern our society (step 2).

As another example, social contract theorists (such as Hobbes and Locke) want to know what justifies the state and what kind of state is justified. For this purpose, they ask us to imagine what would happen in the state of nature (i.e., a scenario where there is no state to regulate persons' conduct) (step 1). The state is justified insofar as life without it is worse than life with it; the kind of state that is justified depends on the kinds of problems that arise in the state of nature (step 2). These are some of the more famous thought experiments among political theorists, but there are many more (Brownlee and Stemplowska 2017; Brun 2018; Miščević 2018).[5]

Thought experiments play a central role in the political theorist's main arguments to justify her most important normative claims. For instance, Rawls's main justification for his theory of justice (known as *justice as fairness*) derives from the purported fact that rational agents will choose it over utilitarianism and other theories of justice in the original position. Similarly, Hobbes's justification for the absolute sovereign crucially depends on the purported fact that without a government with unlimited power the state of nature will descend into a state of universal and perpetual war. Hence, it is critical the political theorist's thought experiment really does generate the specific results she says it does. If it does not, then this would significantly weaken, if not entirely collapse, her normative argument.

[5] Thought experiments are not only used by normative political theorists, of course. They are used in many fields of scholarly inquiry, such as economics (Schabas 2018), physics (Peacock 2018), biology (Schlaepfer and Weber 2018), and more.

But how can we be certain thought experiments generate the specific results the political theorist claims? Many thought experiments are quite complicated. Political theorists typically rely on informal reasoning and intuition when thinking through thought experiments and deriving their logical implications. Yet, informal reasoning and intuition can lead us astray. Many readers will no doubt have experienced cases in which what they initially thought to be intuitively true actually turned out to be false after careful logical scrutiny. Given what is at stake in normative debates, the political theorist's central methodological tool should stand on firmer ground.

To be clear, we are not trying to downplay the value of informal reasoning and intuition. Philosophers of science have distinguished between "discovery" and "justification" of scientific theories/hypotheses. Discovery "concerns the origin, creation, genesis, and invention of scientific theories and hypotheses," while justification "concerns their evaluation, test, defense, success, truth, and confirmation" (Kordig 1978: 110). Intuition and informal reasoning play an important role in the initial discovery stage. For formally inclined social scientists, this usually corresponds to when one initially develops and/or refines various hypotheses, equilibrium concepts, and/or axioms.[6] For political theorists, the discovery stage corresponds to the initial idea and creation of a given thought experiment. However, after a political theorist creates her thought experiment, its subsequent analysis no longer remains at the discovery stage but now enters the justification stage that requires careful logical analysis. Logic and rigorous reasoning, argues Carl R. Kordig, are not essential during the discovery stage. At the latter justification stage, however, "logic is

[6] This can probably be most clearly seen by the label that Cho and Kreps (1987) use to name their proposed refinement of perfect Bayesian equilibrium: "the intuitive criterion."

here essential." (Kordig 1978: 114). We believe this is where formal models can play an important role.

We can think of formal models as precisifications of thought experiments. By precisifying thought experiments via logical and mathematical language, formal models can serve as a bridge that links the discovery phase and the justification phase in a political theorist's normative argument. Note that many thought experiments have a very similar structure to a formal model. There are initial conditions or assumptions (regarding the agents' environment, constraints, rationality, beliefs, desires/preferences, etc.) and there are results or outcomes that are purported to logically follow from these initial conditions or assumptions. The main difference between a formal model and a thought experiment is that unlike an informal thought experiment, all the initial conditions and assumptions in a formal model are (forced to be) stated and defined in precise mathematical and logical language. The main theoretical advantage of this is that once all the initial conditions and assumptions are stated precisely, one can then rely on the rules of deductive logic and mathematical inference to rigorously deduce their logical implications. This allows the political theorist to know with confidence the results of her thought experiment.

Several things might happen once we formally model a political theorist's thought experiment. First, the model might vindicate the theorist's conjecture concerning the outcome of the experiment. This constitutes good news. Second, the model might do the opposite and show that the theorist's conjecture is incorrect. This is precisely what the theorist does not want to happen, but all is not lost if it does. For third and finally, a formal model can show us how the thought experiment can be revised to ensure the theorist's desired conclusion still holds. Such revisions might be unacceptable in that they conflict with other commitments the theorist holds;

if this is true, then the inconsistency between the model and the thought experiment is a serious problem. The revisions might be acceptable, however, and so the model thus helps the political theorist revise and improve her normative theory.

Let's consider an example by looking at Hun Chung's (2015) formal model of Hobbes's state of nature. Recall that Hobbes's primary aim was to justify the state with unlimited powers. Hobbes's justification for such a state entirely derives from the purported fact that without it, the state of nature necessarily results in a state of universal and perpetual war, where life is "solitary, poor, nasty, brutish, and short" (Hobbes 1994: 76). Is Hobbes correct that the state of nature necessarily devolves into such a state?

To begin, Chung offers the following list of features that characterize life in the state of nature, extracted from Hobbes's text:

- C1: people value their lives and universally desire to secure their self-preservation.
- C2: competition over scarce resources (needed for one's survival) inevitably arises.
- C3: people's physical/mental capabilities are roughly equal; so, the weakest has enough strength to kill the strongest, which makes everybody a potential threat to everybody else's self-preservation.
- C4: some people are *vainglorious*: they seek glory and pursue power more than what is minimally needed for their survival.
- C5: there is *uncertainty*; people are unable to tell whether other people are vainglorious or not.

Hobbes's contention is that in such situations, it is rational for everybody to preemptively attack, and as a result, the state of nature necessarily dissolves into a suboptimal state of universal war. To remedy this, we need to establish a government with absolute powers.

Why believe this? There are logical gaps between Hobbes's thought experiment and his normative conclusion. It is not entirely transparent the five conditions from C1 to C5 are jointly sufficient to make it rational for everybody to preemptively attack; whether everybody preemptively attacking necessarily leads to a state of universal war, making war the unique equilibrium of the state of nature; whether a state of universal war would be suboptimal; and whether a government with absolute powers is necessary to escape such a predicament. To examine the validity of such claims, Chung develops the following two-player Bayesian game with two-sided incomplete information to represent Hobbes's state of nature. The game is illustrated in Figure 1.

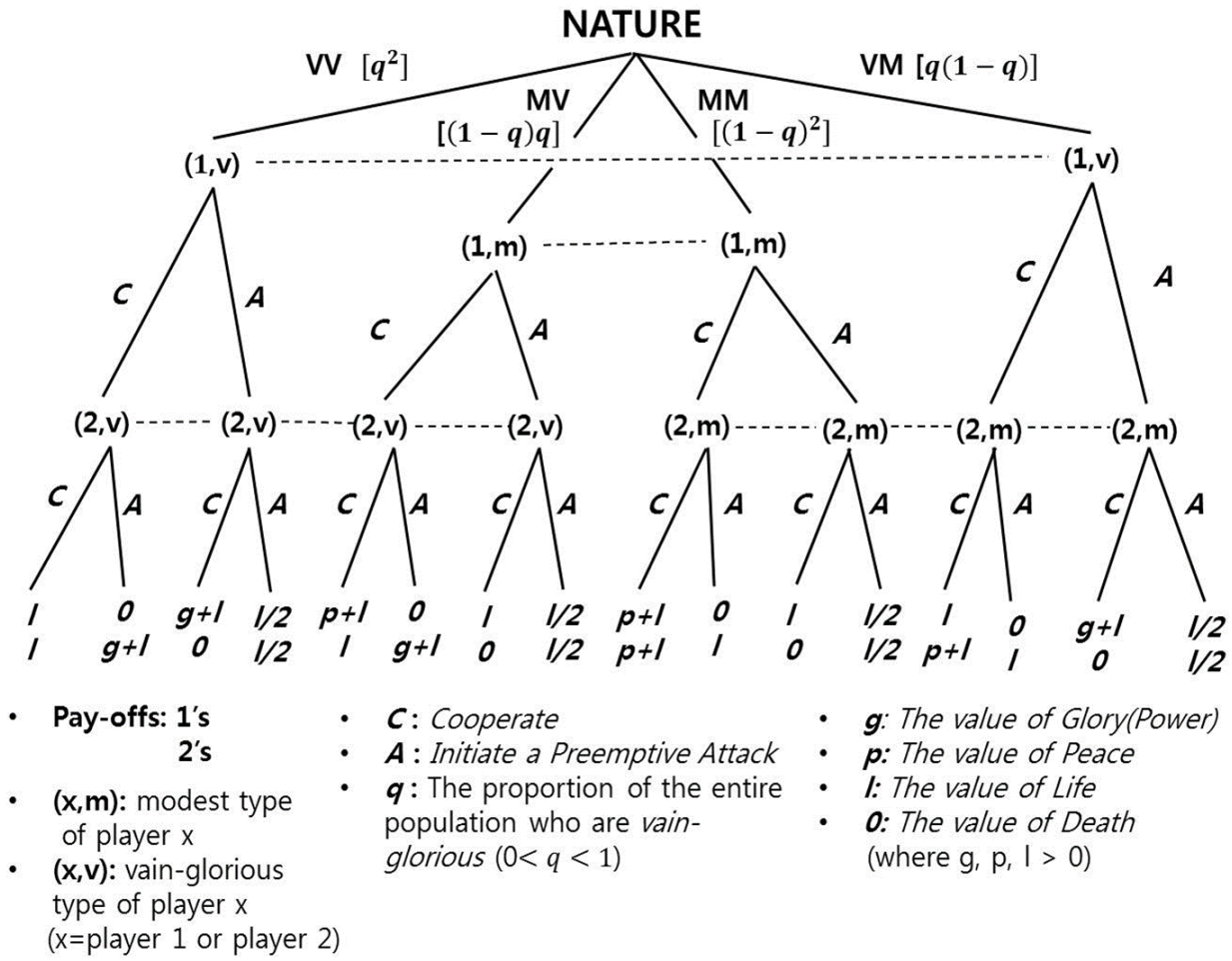

Figure 1: Chung's formal model of Hobbes's state of nature

In the model, each player can either be a modest type or a vainglorious type and must decide whether to cooperate (denoted C) or preemptively attack (denoted A). By mutually cooperating, both players can achieve mutual peace and preserve their lives; by mutually attacking, both players go to war, which (by condition C3) gives each player an equal chance to survive; unilateral cooperation (respectively, unilateral attack) leads to death (respectively, glory). Each player receives a default payoff of $l > 0$ for preserving their lives; the modest types (respectively, vainglorious types) receive an additional payoff of $p > 0$ (respectively, $g > 0$) for achieving peace (respectively, glory.) Here is the timing of the model. Initially, nature makes the

first move and gives a probability distribution over the four possible states of affairs; afterwards, each player, without knowing the other player's type or action, decides whether to cooperate or attack. Payoffs are generated and the game ends. Figure 2 summarizes the main results.

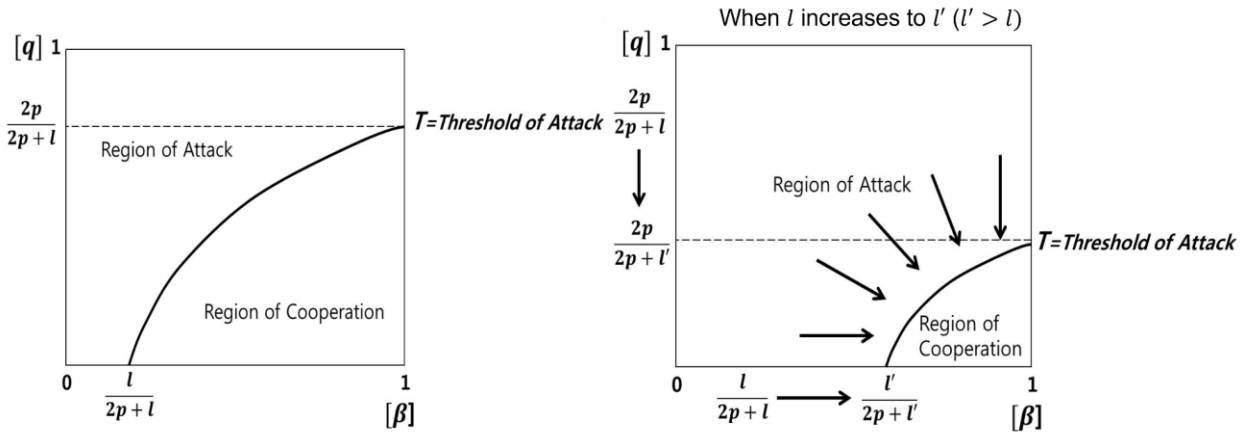

Figure 2: Results of Chung's formal model of Hobbes's state of nature

In the figure, there are two square regions. The horizontal axis of each region represents the probability (β) with which other modest types cooperate, and the vertical axis represents the proportion (q) of vainglorious types in the population. We can see that each square region is demarcated by a line called the "threshold of attack T" that divides the square into two separate regions: the "region of attack" and the "region of cooperation." The interpretation is that for any given value of β, it would be optimal for the modest types to preemptively attack if the proportion q of vainglorious types in the population exceeds the threshold of attack T. We can see that the threshold of attack T is an increasing function of β, implying that as other modest types become more and more likely to cooperate, it takes a higher proportion of vainglorious types in the population to make it optimal to preemptively attack.

What happens when people in Hobbes's state of nature start to attach more value to preserving their own lives (i.e., if the value of l increases)? As Figure 2 shows, this has an effect of *shrinking* the region of cooperation and *expanding* the region of attack. As a consequence, when people start to value their lives arbitrarily highly, the region of cooperation disappears, and the entire region turns into a region of attack. So, the basic result of the model is that for any arbitrarily small proportion of glory-seekers, the state of nature necessarily descends into a state of universal war so long as people value their lives sufficiently highly. This identifies the conditions under which Hobbes's pessimistic conclusion is true.

What do we learn from such a formal model? We learn that there is a sense in which Hobbes was right. Under certain conditions, the state of nature necessarily descends into a state of war. We also learn the main factors that drive this result are *uncertainty* (over other people's types and actions) and the *high valuation of one's own life*. This is inconsistent with what much of the secondary literature says about the causal mechanisms that drive the state of nature into a state of universal war. The standard view says war stems from Hobbes's commitment to psychological egoism, the claim that humans are universally selfish and care only about promoting their power. Not only does this weaken the plausibility of Hobbes's overall argument by basing it on a controversial theory of human nature, but it is also in tension with what Hobbes actually says (we shall have more to say about this in §5 below). Chung's formal model was conducive to correcting the standard interpretation by highlighting the actual mechanisms that drive conflict in the state of nature. It would have been difficult to identify such mechanisms without the help of a formal model.[7]

[7] For other formal models of Hobbes's state of nature, see Vanderschraaf (2006) and Schaefer and Sohn (2022). Vanderschraaf (2006) assumes an infinite type space and shows via computer simulations that a very small number of vainglorious types can *invade* a population consisting mostly of modest types and generate universal war after

3. Possibility and Impossibility

Political theorists often develop normative theories of justice and democracy. In many cases, these theories consist of multiple normative principles. For example, Rawls's justice as fairness consists of the principle of equal basic liberty, the principle of fair equality of opportunity, and the difference principle. Nozick's theory of justice also consists of three principles: just acquisition, transfer, and rectification.

By what criteria should we judge competing theories of justice? This is something political theorists fight about. No one disagrees, however, that principles constituting a theory of justice or democracy must at the very least be *consistent* with one another. For instance, if Nozick's principle of just rectification conflicts with the transfer principle, then that is a problem. Demonstrating the logical consistency of a theory of justice is a necessary but not a sufficient condition for its acceptance. Demonstrating the logical inconsistency of a theory of justice is a decisive reason to reject it.

Problematically, our intuitive judgments about logical consistency are often incorrect. Consider Kenneth Arrow's (2012) impossibility theorem. Arrow demonstrated that several normative criteria we might ideally like a preference aggregation rule to satisfy[8] cannot all be mutually satisfied as long as we have three or more social alternatives. As Amartya Sen notes, this finding "can hardly be anticipated on the basis of common sense or informal reasoning" (Sen

only a couple rounds of repeated play. Schaefer and Sohn (2022) also assume an infinite type space and further assume that the modest types and vainglorious types have "social preferences" (where each type cares about how much material resources they have relative to other players) and show that an arbitrarily small proportion of vainglorious types will spark a chain-reaction that "unravels" the state of nature into a state of war.

[8] This includes Universal Domain, Social Rationality, Weak Pareto, Independence of Irrelevant Alternatives, and No Dictatorship.

2002b: 73). Without formal axiomatization and careful logical analysis, inconsistencies among our firm normative commitments are not easily identifiable by intuitive and informal judgments alone. “Informal insights,” writes Sen, “cannot replace the formal investigations that are needed to examine the congruity and cogency of combinations of values and of apparently plausible demands” (Sen 2002b: 73). Hence, formal tools can be used to determine the consistency of our normative principles.

The value of formal axiomatic analysis is not limited to merely helping us identify the consistency of our normative theories. Once we have understood that several of our firm normative commitments are mutually inconsistent and cannot all be satisfied simultaneously via an impossibility theorem, formal axiomatic analysis can then further help us understand precisely how our normative principles must be revised to render them consistent with one another. Of course, these revisions must be philosophically defensible. Formal tools, however, at the very least point the way forward and highlight options for revision if revision is necessary.

Perhaps the first use of formal tools to examine the logical consistency of a theory of justice was Sen’s (1970) paradox of the Paretian liberal. In this proof Sen demonstrates normative principles that many liberal political theorists purportedly embrace are mutually inconsistent. On our interpretation of the proof, Sen models liberals’ commitment to rights and a personal sphere with the Liberalism axiom, liberals’ commitment to popular sovereignty (a very weak version of it) with the (weak) Pareto axiom, and liberals’ commitment to tolerance (all preference orderings are admissible) with the Unrestricted Domain axiom. The three axioms are inconsistent. Liberals must thus revise their basic normative commitments or explain why Sen’s axioms fail to accurately model them. There have been many attempts at both endeavors (e.g., Gaertner *et al* 1992; Sen 2002a).

A more contemporary example of this use of formal tools among normative political theorists can be found in Tsuyoshi Adachi, Hun Chung, and Takashi Kurihara's recent paper "(The Impossibility of) Deliberation-Consistent Social Choice." There is now a growing consensus among democratic theorists that deliberation and aggregation (i.e., voting) have their own respective virtues, and rather than viewing each as a substitute for the other, we should incorporate both into our democratic processes, where people "first talk, then vote" (Goodin 2008: ch. 6). Suppose then we have a two-stage democratic process in which people deliberate and persuade one another by exchanging reasoned arguments in the first stage, and then vote for the final outcome in the second stage given their post-deliberation preferences. What is the proper normative relationship between the first deliberation stage and the second voting stage? Inspired by the relevant normative political theory literature, Adachi, Chung, and Kurihara propose the following axiom:

> **NNRD (Non-Negative Response toward Democratic Deliberation):** If some individuals, through democratic deliberation in the first stage, change their preferences to align with those of others, then the outcome of the social choice rule in the second voting stage should not leave those who successfully persuaded others through reasoned deliberation worse off than they would have been without deliberation (Adachi *et al* 2023).

NNRD characterizes what the authors believe to be the proper normative relationship between the two democratic stages of deliberation and voting. If NNRD is violated, then people can be made worse-off by persuading others via democratic deliberation. This disincentivizes people to deliberate, which undermines the very purpose of incorporating a separate deliberation stage into our democratic processes. Therefore, it is crucial for any two-stage democratic process involving

both deliberation and voting to satisfy NNRD. Additionally, NNRD is weaker than strategy-proofness and is logically implied by it (Adachi et al. 2023: Proposition 1). Thus, to ensure our democratic processes do not incentivize strategic misrepresentation of preferences, they must at a minimum satisfy NNRD.

It turns out that no social choice rule can simultaneously satisfy NNRD, the Weak Pareto axiom, and the No Universal Vetoer axiom (Adachi et al. 2023: Theorem 1). The Weak Pareto axiom expresses the democratic requirement to respect citizens' unanimous preferences, reflecting a minimal notion of popular sovereignty. The No Universal Vetoer axiom embodies our commitment to political equality by preventing excessive arbitrary decision-making power from being concentrated in a single individual. The authors' impossibility theorem demonstrates that these three normative principles cannot be simultaneously incorporated into a two-stage democratic procedure involving both deliberation and voting.

Then, which normative principle must we adjust, and to what extent, to render the three normative principles mutually consistent? If we relax the Weak Pareto axiom to Top Unanimity — which requires the social choice rule to choose the alternative that is unanimously top-ranked whenever such an alternative exists — we obtain a possibility result (Adachi et al. 2023: Proposition 4). However, relinquishing Weak Pareto means that the democratic process can no longer eliminate dominated and unanimously dis-preferred alternatives identified through democratic deliberation in the first stage. Similarly, if we relax the No Universal Vetoer axiom to No Dictatorship, we achieve another possibility result (Adachi et al. 2023: Proposition 5). The trade-off here is that we must permit a single individual, who, although not a dictator, can always secure either her best or second-best alternative regardless of the preferences of others, which undermines our normative commitment to political equality. Understanding these significant

normative trade-offs with such precision would have been impossible without the aid of formal axiomatic analysis.

4. Institutional Design

Beyond cooking up thought experiments and proposing theories of justice or democracy, political theorists are in the business of designing novel institutional arrangements. For example, some advocate seating legislatures by lottery rather than elections (Guerrero 2014). Others argue we ought to try some form of epistocracy, a system where only the informed can vote (Brennan 2016). Other examples include legislators voting by secret ballot (Kogelmann 2021: ch. 2), tying legislators' compensation to democratically selected performance outcomes (Juarez-Garcia and Schaefer 2022b), allowing career bureaucrats to select legislators (Chan 2014), and more.

Novel institutional arrangements are designed in hopes these institutions better satisfy certain normative criteria when compared to existing institutions. For instance, those in favor of lottery selection believe randomly selecting legislators better achieves the ideals of political equality and descriptive representation when compared to elections. Ideally, examining the effects of institutions is done empirically. This, for instance, is what many do when comparing presidential to parliamentary systems or plurality rule to proportional representation or unicameralism to bicameralism (e.g., Lijphart 2012). By definition, though, there is no empirical record to judge novel institutions against. In these cases, how do we get a sense of how novel institutions will work in practice? Political theorists typically speculate about the incentive effects of novel institutions in an informal manner.

Formal models can make the analysis of untried institutions more rigorous. In fact, formal models provide "an explicit and systematic methodology for studying the effects of institutions" (Weingast 1996: 168). Institutions are modeled formally "via their effects on the set of actions available to each individual, on the sequence of actions, and on the structure of information available to each decision-maker" (Weingast 1996: 169). By creating formal models of novel institutions, political theorists have firmer ground to stand on when thinking about how their proposals might work in practice. Social scientists use formal models to study existing and past political institutions; there is no reason these tools cannot be used to study possible institutions as well.

Using formal tools to design institutions is not a new idea. Consider *market design*, a subfield of economics (Roth 2007; 2015). Relying mostly on game theory (but some experimental work as well) market designers devise incentive systems to ensure efficient outcomes in markets that are susceptible to failure. They do this by designing institutions to ensure sufficient market thickness, incentivize honest information revelation, manage congestion, and so on. Market designers have used formal tools to help design kidney exchanges, medical professional matching programs, school choice programs, radio spectrum auctions, and more. For instance, clearing houses that match medical professionals to employers use a variant of the algorithm developed by game theorists David Gale and Lloyd Shapely (1962). Using formal tools to think through the design of institutions is not new.

This role for formal models is least used among formally inclined political theorists, which means there is low-hanging fruit to grab. There are some examples, though. For instance, political theorists inspired by Locke debate the best institutional rules for regulating the appropriation of unclaimed natural resources. *Right libertarians* believe there should be no or

few constraints on how much first appropriators may take. On the other hand, *left libertarians* believe persons should be constrained in how much they take leaving, in Locke's words, "enough and as good" for others (Locke 1980: 21). Debate between right and left libertarians has been going on for some time with little progress made. These debates are often over the effects of the different systems of appropriation. Given that we have no good data to adjudicate this dispute, who is right?

Brian Kogelmann and Benjamin Ogden (2018) develop a formal model to shed light on this question. Using a general equilibrium framework, they compare an institutional rule where households may appropriate as much land as they like to an institutional rule where households can only appropriate a prespecified amount of land. Once appropriated, land can be invested in (which increases future productivity), members of other households can be hired to work the land, and trade of goods produced from the land commences. With this framework, Kogelmann and Ogden show that the right libertarian appropriation rule will almost always dominate the left libertarian rule from a welfare perspective. This is because early increases in productivity under a right libertarian rule (more land will be invested in earlier) trump welfare losses that later generations face from being unable to appropriate land. The result is consistent with philosopher David Schmidtz's conjecture that "original appropriation benefits latecomers far more than it benefits original appropriators. Original appropriation is a cornucopia of wealth, but mainly for latecomers" (Schmidtz 2008: 196). With no empirical record to adjudicate the dispute between two different systems of natural resource appropriation, the model makes some progress in an otherwise stagnant debate.

5. Evaluating Normative Models

We have highlighted three roles for formal models in normative political theory. How do we appraise these models, though? To put it another way: what separates good normative political theory models from bad ones?

When it comes to evaluating models generally—be they normative or not—we follow Kevin A. Clarke and David M. Primo who hold that "models are purpose-relative and that model assessment cannot take place without understanding the purpose for which the model was designed and used … we need to ask whether a model is useful for the purpose for which it was intended" (Clarke and Primo 2012: 61). For example, some social scientists build models to explain descriptive patterns. If this is true, then the assumptions of the model should correspond to the trade-offs faced by real-world actors (Paine and Tyson 2020). As another example, some models are built to isolate causal mechanisms. These models should be judged according to their parsimony, for parsimony facilitates conceptual clarity (Paine and Tyson 2020).

Like all models, formal normative models should be judged according to their purpose. We have highlighted three uses of formal models among political theorists, but all three uses have one thing in common: they allow us to substitute political theorists' informal conjectures with more rigorous inferences. Instead of relying on Hobbes's speculation concerning what the state of nature is like, for instance, we can build a formal model of it. Thus, a formal model of a political theorist's normative theory should be faithful to the intentions of the theorist. The assumptions of the model should be grounded in the text. If we build a model of Hobbes's state of nature, to continue the example, then every assumption should be supported by what Hobbes says in *Leviathan* and other writings.

Let's apply this evaluative framework by returning to formal models of Hobbes's state of nature, as discussed in §2 above. For some time now, Hobbes scholars have used the tools of

game theory to better understand Hobbes's assertion that the state of nature results in a state of war. Some modeled the state of nature as a one-shot Prisoners' Dilemma (e.g., Gauthier 1969: 79-80; Rawls 1971: 269), others as an iterated Prisoners' Dilemma (e.g., Hampton 1986: 75-89; Kavka 1986: 129-136). Still others have argued the state of nature is best modeled as an assurance game (e.g., Moehler 2009).

Problematically, using these models to interpret and understand the famous thought experiment is in tension with Hobbes's text. These models assume players' ordinal preference rankings are identical. In the Prisoners' Dilemma both players most prefer to defect while the other cooperates; in the assurance game, both players most prefer mutual cooperation, but unilateral cooperation results in their worst-case outcome, which can lead to mutual defection. Hobbes is very clear, however, that preferences in the state of nature are heterogenous. Consider the following passage:

> In the state of nature there is in all men a will to do harm, but *not for the same reason or with equal culpability.* One man practices the equality of nature, and allows others everything which he allows himself; this is the mark of a *modest man* . . . Another, supposing himself superior to others, wants to be allowed everything, and demands more honour for himself than others have; that is the sign of an aggressive character. In his case, the will to do harm derives from *vainglory* and over-valuation of his own strength (Hobbes 2016: 26; emphasis ours).

In this passage, Hobbes distinguishes between two types of people: the modest and the vainglorious. The modest types are those who curb their self-interest and conditionally cooperate with other cooperators if doing so allows them to achieve long-lasting mutual peace. By contrast, the vainglorious types will always take advantage of others' cooperative behavior and

preemptively attack to conquer and forever increase their own power. For them, preemptive attack is a strictly dominant strategy.

The point here is that the early formal models of Hobbes's state of nature were flawed because their assumptions were inconsistent with the text. If the goal is to model Hobbes's state of nature, assumptions must better align with what we find in *Leviathan* and other writings. The more recent models of Hobbes's state of nature avoid this shortcoming. For instance, Chung's model—which we discussed in §2 above—explicitly assumes there are two different types of players; indeed, Chung shows that it is uncertainty over other players' type that drives conflict. Chung's model of the state of nature along with other more recent ones (e.g., Vanderschraaf 2006; Schaefer and Sohn 2022) are superior to the prior generation of models because their assumptions better align with the original text.

6. Changing Political Theory for the Worse?

We believe political theorists should rely on formal models to a greater extent. An objection to this imperative worries that increased reliance on formal theory among political theorists changes the subfield. Among philosophers of science, it is well-known that many motivational factors influence where scholars allocate their research efforts (Kitcher 1993: ch. 8; D'Agostino 2010: ch. 4-5). One's research paradigm—such as rational choice theory—plays a role. As Fred D'Agostino writes: "the availability of a paradigm, and enquirers' principled commitment to it, has a tendency to *channel* research and to raise barriers to new ideas" (D'Agostino 2010: 50). The concern here is that if political theorists embrace formal models, it will change where they focus their time and attention.

Why would embracing formal models influence where political theorists allocate their research efforts? There are many reasons. First, not all thought experiments are amenable to formal reconstruction. Only those that involve complicated choice problems are worth modeling. We do not see how formal models can illuminate the infamous trolley problem thought experiment, for example. Moreover, if a theory of justice or democracy only consists of one normative principle, then there is nothing to scrutinize in terms of its logical consistency. Furthermore, some political theorists are uninterested in the institutional implications of their normative principles (e.g., Cohen 2008: ch. 6); formally inclined political theorists will pay less attention to their work. Finally, if a political theorist rejects the basic tenets of rational choice theory, then a rational choice model cannot shed light on her work.

We do not deny greater embrace of formal models will change where political theorists allocate their time and attention. Indeed, among the current formally inclined political theorists, there is much focus on social contract theories (Hobbes, Locke, Rousseau, and Rawls) precisely because this kind of normative work lends itself to game theoretic models. That said, we are not too concerned with political theorists shifting their research interests in response to adopting formal theory, for two reasons.

First, how political theorists *currently* allocate their research efforts is tainted by existing paradigms. We should not pretend political theorists today are uninfluenced by their paradigms, choosing to research whatever they think yields the greatest social benefit. All scholars exist within paradigms, which colors the research questions they pursue. So, it is no objection to say embracing formal theory will influence where political theorists allocate their time and attention, for the current dominant research methods also exert influence. For this objection to land, the critic of formal theory would need to go further and argue that political theorists influenced by

formal theory would choose substantively *worse* research topics than they currently do. We see no reason to believe this.

Second, we do not believe all political theorists should use formal models (though we do think they should all have a working familiarity with them, a claim we argue for in the next section). We are claiming that (*i*) formal theory has immense value for normative political theory and (*ii*) the number of political theorists who use formal models in their research is currently suboptimal. Saying the number is suboptimal does not imply that more will always be better. Disciplines should be characterized by multiple research methods. Research on the division of cognitive labor suggests diverse scholarly communities outperform homogenous ones (Muldoon 2013). Political theory (in our estimation) needs more people who work with formal models, but it still needs scholars who work on the history of thought, analytic political philosophy, critical theory, comparative political theory, and more. As we noted, some normative questions may elude being stated in formal language; we still think it is a good thing for political theorists to research these topics without formal tools.

7. Does Political Theory Belong in Political Science?

Political theory occupies a tenuous place in political science. It sometimes feels unwelcome. These feelings are not new. A 1957 article published in the *American Political Science Review* on political theory's place in the discipline begins: "Among political scientists, even among political theorists, there is a widespread conviction that political theory has entered upon a time of troubles" (Smith 1957: 734). Not infrequently does one hear political scientists question whether political theory belongs in the discipline. Whether it does or not depends on

two things: first, how political science as an academic discipline is defined, and second, the characteristics of political theory research. To end, we shall argue that greater embrace of formal models can only strengthen political theory's place in the discipline.

To begin, how are the boundaries of political science defined? A broad definition says political science includes all research that is focused on politics (Corbett 2011). If this is how political science is defined, then political theory belongs in political science, because political theorists do research on politics. This definition of political science is problematic, however. This is because defining academic disciplines *solely* in terms of subject matter leads to counterintuitive implications. Astrophysicists and astrologists both study celestial bodies, for instance, but they do not belong in the same department. Chemists and alchemists study similar phenomena, but they also do not belong in the same department. No doubt an academic discipline is *partly* defined in terms of its subject matter, but demarcating disciplinary boundaries must consist of more than that.

Andrew Rehfeld takes a stab at what that "more" consists of. On his view, for research to count as political science, it must take political phenomena as its subject, it must assume there is an observer-independent world about which its aim is to discover facts, and its claims must be falsifiable (Rehfeld 2010: 472). If this is how political science is defined, then some political theory research will count as political science, but some will not (Rehfeld 2010: 474-479). We agree with Rehfeld that academic disciplines must be defined in terms of subject matter plus something else, but we disagree with his proposed something else, for two reasons. First, as philosophers of science have long known, falsifiability as a demarcation criterion for genuine science is riddled with difficulties, especially when applied to the social sciences (Clarke and Primo 2012: 33-41). Second, many intuitively important aspects of political science—such as

many celebrated game theory and social choice models—are analytical results not subject to falsification.

In our view, an academic discipline is defined in terms of its subject matter along with its methodological toolkit. Astrophysicists and astrologists are interested in a similar subject matter, but use very different methodological tools, which is why they don't belong in the same department. Practitioners in a discipline need not use *every* tool in the disciplinary toolkit to do their research—some disciplines' methodological toolkits are quite large—but they should (*i*) use *only* tools in the disciplinary toolkit to do their research, and (*ii*) should have at least a working familiarity with all the tools. It is not our job to say what political science's methodological toolkit is. At the very least, though, it includes formal modeling based on rational choice.

Does political theory count as political science according to this way of understanding the discipline? That depends on the tools that ultimately constitute political science's methodological toolkit as well as how political theorists go about their work, which are questions we will not answer. What we can say, however, is that engaging more with formal models can only help political theory's case. Since rational choice formal modeling is part of the political science toolkit, political theorists who deploy formal models in their research are definitely part of the discipline. Moreover, even those who do not use formal models in their research should at least have a working familiarity with these tools, for all practitioners in a discipline should be familiar with its basic tools. If more political theorists embrace formal models, then we believe the subfield would have a stronger foothold in the discipline which, in our view, would be a good thing.

Works Cited

Adachi, Tsuyoshi, Hun Chung, and Takashi Kurihara. 2023. "(The Impossibility of) Deliberation-Consistent Social Choice." Forthcoming in *American Journal of Political Science*.

Amadae, S. M. and Bruce Bueno de Mesquita. 1999. "The Rochester School: The Origins of Positive Political Theory." *Annual Review of Political Science* 2: 269-295

Arrow, Kenneth. 2012. *Social Choice and Individual Values*. New Haven: Yale University Press.

Barrett, Jacob. 2020. "Is Maximin Egalitarian?" *Synthese* 197: 818-837.

Brennan, Jason. 2016. *Against Democracy*. Princeton: Princeton University Press.

Braham, Matthew and Martin van Hees. 2014. "The Impossibility of Pure Libertarianism." *Journal of Philosophy* 111: 420-436.

Brownlee, Kimberly and Zofia Stemplowska. 2017. "Thought Experiments." In *Methods in Analytical Political Theory,* edited by Adrian Blau: 21-45. Cambridge: Cambridge University Press.

Brun, Georg. 2018. "Thought Experiments in Ethics." In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown: 195-210. New York: Routledge.

Bruner, Justin. 2015. "Diversity, Tolerance, and the Social Contract." *Politics, Philosophy & Economics* 14: 429-448.

Bruner, Justin. 2020. "Locke, Nozick, and the State of Nature." *Philosophical Studies* 177: 705-726.

Chan, Joseph. 2014. *Confucian Perfectionism: A Political Philosophy for Modern Times*. Princeton: Princeton University Press.

Cho, In-Koo and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102: 179-222.

Chung, Hun. 2015. "Hobbes's State of Nature: A Modern Bayesian Game-Theoretic Analysis." *Journal of the American Philosophical Association* 1: 485-508.

Chung, Hun. 2019. "The Impossibility of Liberal Rights in a Diverse World." *Economics & Philosophy* 35: 1-27.

Chung, Hun. 2020. "Rawls's Self-Defeat: A Formal Analysis." *Erkenntnis* 85: 1169-1197.

Chung, Hun. 2022a. "Locke's State of Nature and its Epistemic Deficit: A Game-Theoretic Analysis." *Synthese* 200: 147.

Chung, Hun. 2022b. "When Utilitarianism Dominates Justice as Fairness: An Economic Defense of Utilitarianism from the Original Position." *Economics & Philosophy* (FirstView).

Chung, Hun and John Duggan. 2020. "A Formal Theory of Democratic Deliberation." *American Political Science Review* 114: 14-35.

Chung, Hun and Brian Kogelmann. 2020. "Diversity and Rights: a Social Choice-Theoretic Analysis of the Possibility of Public Reason." *Synthese* 197: 839-865.

Clarke, Kevin A. and David M. Primo. 2012. *A Model Discipline: Political Science and the Logic of Representations*. Oxford: Oxford University Press.

Cohen, G.A. 2008. *Rescuing Justice and Equality*. Cambridge: Harvard University Press.

Corbett, Ross J. 2011. “Political Theory within Political Science.” *PS: Political Science and Politics* 44: 565-570.f

D’Agostino, Fred. 2010. *Naturalizing Epistemology: Thomas Kuhn and the ‘Essential Tension.’* New York: Palgrave Macmillan.

Diermeier, Daniel. 2015. “Positive Political Theory.” In *The Encyclopedia of Political Thought,* edited by Michael Gibbons: 1-9. John Wiley & Sons. DOI: 10.1002/9781118474396.wbept0810.

Fiorina, Morris P. 1975. “Formal Models in Political Science.” *American Journal of Political Science* 19: 133-159.

Forbes, Donald H. 2004. “Positive Political Theory.” In *Handbook of Political Theory,* edited by Gerald F. Gaus and Chandran Kukathas: 57-72. London: SAGE Publications.

Gaertner, Wulf, Prasanta K. Pattanaik, and Kotaro Suzumura. 1992. “Individual Rights Revisited.” *Economica* 59: 161-177.

Gale, David and Lloyd Shapeley. 1962. “College Admissions and the Stability of Marriage.” *The American Mathematical Monthly* 69: 9-15.

Gauthier, David. 1969. *The Logic of Leviathan*. Oxford: Oxford University Press.

Goodin, Robert E. 2008. *Innovating Democracy: Democratic Theory and Practice After the Deliberative Turn*. Oxford: Oxford University Press.

Guerrero, Alexander. 2014. “Against Elections: The Lottocratic Alternative.” *Philosophy & Public Affairs* 42: 135-178.

Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.

Hobbes, Thomas. 1994. *Leviathan*, edited by Edwin Curley. Indianapolis: Hackett Publishing.

Hobbes, Thomas. 2016. *On the Citizen*, edited by Richard Tuck and Michael Silverthorne. Cambridge: Cambridge University Press.

Ingham, Sean. 2019. *Rule by Multiple Majorities: A New Theory of Popular Control*. Cambridge: Cambridge University Press.

Ingham, Sean and Frank Lovett. 2019. "Republican Freedom, Popular Control, and Collective Action." *American Journal of Political Science* 63: 774-787.

Ingham, Sean and Frank Lovett. 2022. "Domination and Democratic Legislation." *Politics, Philosophy & Economics* 21: 97-121.

Juarez-Garcia, Mario and Alexander Schaefer. 2022a. "Exit and Isolation: Rousseau's State of Nature." *Synthese* 200: 252.

Juarez-Garcia, Mario and Alexander Schaefer. 2022b. "Public Servants." *Journal of Moral Philosophy* (Online First).

Johnson, James. 2014. "Models Among the Political Theorists." *American Journal of Political Science* 58: 547-560.

Kavka, Gregory S. 1986. *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Knight, Jack and James Johnson. 2015. “On Attempts to Gerrymander ‘Positive’ and ‘Normative’ Political Theory: Six Theses.” *The Good Society* 24: 30-48

Kogelmann, Brian. 2019. “Public Reason’s Chaos Theorem.” *Episteme* 16: 200-219.

Kogelmann, Brian. 2021. *Secret Government: The Pathologies of Publicity*. Cambridge: Cambridge University Press.

Kogelmann, Brian and Benjamin Ogden. 2018. “Enough and as Good: A Formal Model of Lockean First Appropriation.” *American Journal of Political Science* 62: 682-694.

Kordig, Carl R. 1978. “Discovery and Justification.” *Philosophy of Science* 45: 110-117.

Lijphart, Arend. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven: Yale University Press.

Locke, John. 1980. *Second Treatise of Government*, edited by C.B. Macpherson. Indianapolis: Hackett Publishing.

Mershon, Carol and Olga Shvetsova. 2019. *Formal Modeling in Social Science*. Ann Arbor: University of Michigan Press.

Miščević, Nenad. 2018. “Thought Experiments in Political Philosophy.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown: 153-170. New York: Routledge.

Moehler, Michael. 2009. “Why Hobbes’s State of Nature is Best Modeled by an Assurance Game.” *Utilitas* 21: 297-326.

Moreno-Ternero, Juan D. and John E. Roemer. 2008. “The Veil of Ignorance Violates Priority.” *Economics & Philosophy* 24: 233-257.

Motchoulski, Alexander. 2021. “Adjudicating Distributive Disagreement.” *Synthese* 198: 5977-6008.

Muldoon, Ryan. 2013. “Diversity and the Division of Cognitive Labor.” *Philosophy Compass* 8: 117-125.

O’Connor, Cailin. 2019. *The Origins of Unfairness: Social Categories and Cultural Evolution*. Oxford: Oxford University Press.

Paine, Jack and Scott A. Tyson. 2020. “Uses and Abuses of Formal Models in Political Science.” In *The SAGE Handbook of Political Science*, edited by Dirk Berg-Schlosser, Bertrand Badie, and Leonardo Morlino: 188-202. London: SAGE Publications.

Page, Scott E. 2021. *The Model Thinker: What You Need to Know to Make Data Work for You*. New York: Basic Books.

Patty, John W. and Elizabeth Maggie Penn. 2014. *Social Choice and Legitimacy: The Possibility of Impossibilities*. Cambridge: Cambridge University Press.

Peacock, Kent A. 2018. “Happiest Thoughts: Great Thought Experiments of Modern Physics.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown: 211-242. New York: Routledge.

Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

Rehfeld, Andrew. 2010. “Offensive Political Theory.” *Perspectives on Politics* 8: 465-486.

Roemer, John E. 1996. *Theories of Distributive Justice*. Cambridge: Harvard University Press.

Roemer, John E. 2004. “Eclectic Distributional Ethics.” *Politics, Philosophy & Economics* 3: 267-281.

Roth, Alvin E. 2007. “The Art of Designing Markets.” *Harvard Business Review* October Issue: 118-126.

Roth, Alvin E. 2015. *Who Gets What–and Why: The New Economics of Matchmaking and Market Design*. Boston: Houghton Mifflin Harcourt.

Schabas, Margaret. 2018. “Thought Experiments in Economics.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown: 171-182. New York: Routledge.

Schaefer, Alexander. 2021. “Rationality, Uncertainty, and Unanimity: An Epistemic Critique of Contractarianism.” *Economics & Philosophy* 37: 82-117.

Schaefer, Alexander and Jin-Yeong Sohn. 2022. “Unraveling into War: Trust and Social Preferences in Hobbes’s State of Nature.” *Economics & Philosophy* 38: 171-205.

Schlaepfer, Guillaume and Marcel Wejuber. 2018. “Thought Experiments in Biology.” In *The Routledge Companion to Thought Experiments*, edited by Michael T. Stuart, Yiftach Fehige, and James Robert Brown: 2th43-256. New York: Routledge.

Schmidtz, David. 2008. *Person, Polis, Planet: Essays in Applied Philosophy*. Oxford: Oxford University Press.

Sen, Amartya. 1970. “The Impossibility of a Paretian Liberal.” *Journal of Political Economy* 78: 1252-157.

Sen, Amartya. 2002a. “Liberty and Social Choice.” In *Rationality and Freedom*: 381-407. Cambridge: Harvard University Press.

Sen, Amartya. 2002b. “The Possibility of Social Choice.” In *Rationality and Freedom*: 65-120. Cambridge: Harvard University Press.

Smith, David G. 1957. “Political Science and Political Theory.” *American Political Science Review* 51: 734-746.

Tungodden, Bertil and Peter Vallentyne. 2005. “On the Possibility of Paretian Egalitarianism.” *Journal of Philosophy* 102: 126-154.

Vanderschraaf, Peter. 2006. “War or Peace? A Dynamical Analysis of Anarchy.” *Economics & Philosophy* 22: 243-279.

Vanderschraaf, Peter. 2010. “The Invisible Foole.” *Philosophical Studies* 147: 37.

Vanderschraaf, Peter. 2019. *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. Oxford: Oxford University Press.

Weingast, Barry. 1996. “Political Institutions: Rational Choice Perspectives.” In *A New Handbook of Political Science*, edited by Robert E. Goodin and Hans-Dieter Klingerman: 167-190. Oxford: Oxford University Press.