Making fair comparisons in political theory

Sean Ingham and David Wiens

Abstract. Normative political theorists frequently compare hypothetical scenarios for the purpose of identifying reasons to prefer one kind of institution to alternatives. We examine three types of "unfair" comparisons and the reasoning errors associated with each. A theorist makes an *obscure comparison* when one (or more) of the alternatives under consideration is underspecified; a theorist makes a *mismatched comparison* when they fail to hold fixed the relevant contextual factors while comparing alternatives; and a theorist makes an *irrelevant comparison* when they compare alternatives assuming contextual factors that differ in important respects from those they "should" assume given their theoretical aims. We then introduce the notion of a modeling mindset and show how this mindset can help theorists detect and avoid the three types of error. We conclude with a reconstruction of Cohen's (2009) camping trip thought experiment to illustrate the approach.

Normative political theorists frequently compare hypothetical scenarios for the purpose of examining our reasons to prefer one kind of social and political arrangement over alternatives. For example, a theorist might compare a democratic society to non-democratic societies to expose the advantages of democracy (e.g., Bagg 2018). (We illustrate the breadth of this practice with several additional examples below.) Political scientists have spent considerable effort developing best practices for comparative reasoning in empirical research, paying special attention to causal inference (among others, see Collier and Munck 2022; Egami and Hartman 2023; McDermott 2011; Morgan and Winship 2015; Sartori 1991). While political theorists might look to this literature for suggestive analogies (see, e.g., Dowding 2022), the methodological problems that arise in normative comparative reasoning deserve their own treatment. Normative theorists do not use comparative reasoning to answer questions about the causal effects of institutions or policies, nor any of the other kinds of questions empirical social scientists target. So there is no reason to expect methodological concepts developed for comparative reasoning in empirical social science to have much purchase for normative theory—perhaps careful inquiry

To appear in *American Journal of Political Science*.

would reveal they do, but that would be a genuine and unexpected discovery.[1] Yet while normative political theorists cannot just grab existing methodological insights off the empiricists' shelf, they should follow their colleagues in systematically reflecting on how their distinctive kind of comparative reasoning about social and political arrangements can fail and what to do to guard against failure. We take a first step in this direction.

We proceed in two stages. In the first, we examine three types of "unfair" comparisons and the reasoning errors associated with each. Briefly: a theorist makes an *obscure comparison* when one (or more) of the alternatives under consideration is underspecified; a theorist makes a *mismatched comparison* when they fail to hold fixed the relevant contextual factors while comparing alternatives; and a theorist makes an *irrelevant comparison* when they compare alternatives assuming contextual factors that differ in important respects from those they "should" assume given their theoretical aims.

In the second stage, we show how a modeling mindset helps theorists detect and avoid the three types of error. A *modeling mindset* is a set of dispositions: a disposition to clearly specify all the important aspects of the alternatives under comparison, which helps one avoid obscure comparisons; a disposition to think carefully about which aspects of the alternatives under comparison are being held constant and which are permitted to vary, which helps one avoid mismatched comparisons; and, from a recognition that modeling involves deliberate abstraction from certain details and stylized representation of others, a disposition to entertain questions about the relevance of a particular comparison for one's theoretical purpose, which helps one avoid irrelevant comparisons.

We conclude with an extended example to show the modeling mindset in action. Critics have charged G. A. Cohen's (2009) camping trip thought experiment with all three types of unfair comparison (e.g., Brennan 2014; Krause 2010; Ronzoni 2011; Schmidtz 2011). We develop a revised version of the thought experiment to demonstrate how a modeling mindset can help Cohen avoid these charges.

To avoid misunderstanding, we register a caveat at the outset. Talk of a "modeling mindset" might cue some to expect us to advise political theorists to use *formal* models more often. This is not our aim. A model need not be stated in mathematical terms. It is nothing more than a stylized description of a hypothetical scenario, constructed for the purpose of rendering certain abstract ideas concrete. So understood, models

---

1 As we will see, there are some suggestive analogies and parallels to be drawn, as others have argued (e.g., Dowding 2022). But our point is that their appropriateness cannot be assumed at the outset; one cannot simply draw on existing insights from empirical methodology, which concern causal inferences, induction, and the like, and assume they apply to a domain of inquiry that concerns entirely different categories of propositions.

are ubiquitous in political theory (see Johnson 2014; Wolin 2004, chap. 1), and their use is integral to the kinds of comparisons we take as our focus. We intend to explain the benefits of cultivating a more self-conscious and structured approach to the use of models — formal or informal — in political theory.

1.   UNFAIR COMPARISONS: A TYPOLOGY

Scholars have noted various ways in which normative theorists' comparisons of social and political arrangements can go awry. These critiques have been presented in piecemeal fashion, each one allegedly showing how a specific comparison made by a specific theorist is "unfair" in some specific sense (e.g., Brennan 2014; Freiman 2017; Landa and Pevnick 2021; Valentini 2011; Ypi 2010). Piecemeal critiques have two limitations. First, they obscure the generality of the alleged problem, precluding a general explication of where a theorist's reasoning has gone awry and the conditions that must obtain for the alleged mistake to occur. Second, they obscure important differences between distinct ways in which a comparison can be invalid. Without a typology of unfair comparisons, theorists are liable to make three types of *diagnostic error*: they might fail to detect when a particular comparative mode of reasoning is invalid; they might conflate distinct modes of invalid reasoning; or they might charge a theorist with invalid comparative reasoning when, in fact, the charge does not apply. The last two of these can lead to erroneous prescriptions for correcting the mistaken reasoning.

   A general typology of unfair comparisons helps to address the limitations of piecemeal critiques, while bringing two additional advantages. First, by abstracting from particular examples and formulating general categories of unfair comparisons, we can more easily identify parallels with comparative reasoning in domains of inquiry besides normative political theory. We note an example below. Second, and relatedly, it paves the way for constructive thinking about how to detect and avoid unfair comparisons. Once we have presented our typology, we will see that detecting and avoiding unfair comparisons is a matter of asking oneself a series of questions, which form a natural progression:

- Are the key features of the alternatives clearly and consistently specified?

- When comparing the alternatives, is it clear which of their aspects are being held fixed and which are allowed to vary?

- Are the alternatives under comparison constructed in a way that allows one to infer conclusions that are relevant given the stated theoretical aims?

As we will demonstrate later, these are exactly the kinds of questions that researchers with a modeling mindset are disposed to ask both when assessing others' comparisons and when presenting their own. Thus, a general typology of unfair comparisons allows us to see clearly how cultivating a modeling mindset induces theorists to adopt a familiar set of best practices for avoiding and detecting common errors of comparative reasoning.

Some terminology will aid our exposition. A *model* is a simplified description of a state of affairs. As an example, consider the following description of a hiring process:

Hiring Committee. A committee of three people is tasked with hiring one of two job applicants, *A* and *B*. Two committee members prefer *A*, while the third prefers *B*. The committee decides whom to hire using majority rule. After taking the vote, the committee hires *A*.

For our purposes, a model of a social and/or political scenario can be decomposed into two components: a structure and a context. A *structure* is a simplified description of a formal institutional scheme or an informal social practice or the like; it is that which determines the actions available to individuals and the manner in which different combinations of choices produce outcomes. The majority voting rule is the structural component of *Hiring Committee*. We think of a model's structure as a mechanism for turning inputs into outcomes. For example, under majority rule, a shared preference for candidate *A* among at least two committee members (input) results in the hiring of candidate *A* (output). In all the examples we will focus on, the structural components of the models are institutional arrangements, like simple majority rule in *Hiring Committee*, so we will refer to them as *institutions*, *institutional arrangements*, or just *arrangements* for short.

A *context* is a simplified description of a range of factors that are extraneous to the model's structure but that can nonetheless influence its operation and the outcomes it produces. The contextual factors in *Hiring Committee* include the identities of the applicants and the preferences of the committee members. The outcome in our example — the hiring decision — depends on our specification of these parameters. For example, the committee would have hired *B* instead if we had assumed that two committee members prefer *B* to *A*. More generally, examples of contextual factors that appear in theorists' models include: the quantity and quality of available natural or technological resources; the distribution of cognitive abilities, preferences, or motivational drives within a population; and so on.

We will occasionally use some formal notation to express general points clearly. We will use *A* and *A'* to denote generic arrangements and *C* and *C'* to denote generic sets

of contextual factors; we will use $\langle A, C \rangle$ to denote a generic model with A and C as its components.

To fix the scope of our attention, we will distinguish between two kinds of comparisons theorists might make. To give some intuition for what we have in mind, let's imagine that we are comparing *Hiring Committee* with an alternative model — call it *Hiring Dictator* — that is identical in every respect except for the decision rule: one committee member is appointed the chair and the chair unilaterally decides whom to hire after consulting the other members. In one type of comparison between these two models, we limit our attention to features that constitute the arrangements under comparison. For example, we might compare the two decision rules with respect to the manner in which they distribute formal voting power. Call this type of comparison a *constitutive comparison*. Alternatively, we might compare the arrangements in these two scenarios with respect not only to their constitutive features but also certain features they exhibit in some contexts but not others. For example, we might compare the decision rules in *Hiring Committee* and *Hiring Dictator* not only with respect to the manner in which they distribute voting power, but also with respect to their tendencies to select high-quality candidates. This latter feature depends not only on constitutive features of the decision rule, but also on the outcomes that result across a range of contexts, which vary along several dimensions, such as the distribution of committee member preferences. Call this second type of comparison a *contingent comparison*. We will focus our attention on several ways in which theorists' contingent comparisons can be considered "unfair."

1.1.   Obscure comparisons   A theorist makes an *obscure comparison* when key aspects of a model scenario under comparison are inadequately specified — perhaps the features constituting an arrangement are unclear or perhaps the context within which the arrangements are assumed to operate are underspecified. For example, consider how Jason Brennan specifies epistocracy: "A system is epistocratic to the extent that greater knowledge and the good faith to act on this knowledge are de jure, legal prerequisites for holding power or are legal grounds for being granted greater power through law" (Brennan 2018, 54). This is a description not of a concrete epistocratic suffrage rule but of the objective at which such a rule should aim. Yet Brennan compares "epistocracy," so described, not with a comparably vague description of democracy (e.g., "rule by the people"), but with actually existing democracies and their actual suffrage rules. Since knowledge and good faith are latent traits, a well-defined epistocratic suffrage rule must specify a test for determining the presence of adequate knowledge and good faith on the basis of some set of observable traits, comparable (in its concreteness) to the voter eligibility requirements used in the actual democracies Brennan criticizes. Absent an

institutional operationalization of an epistocratic suffrage rule, any comparison between actual democracies and hypothetical epistocracies is obscure (Ingham and Wiens 2021, 327).[2]

A theorist commits the *fallacy of obscure comparison* when they infer that one arrangement is better (in some respect) than an alternative on the basis of an obscure comparison. This kind of inference is invalid because the conclusion might hold under some but not all reasonable disambiguations of the ambiguously specified alternative. To continue with the example of epistocracy, disenfranchising citizens who demonstrate ignorance of basic facts about climate change would produce a different electorate than disenfranchising citizens who fail a graduate-level microeconomics exam. The two electorates would differ not only with respect to the levels of knowledge about climate change and microeconomics but also with respect to whatever underlying attributes influence acquisition of such knowledge, attributes that might also, by independent causal pathways, influence voting behavior (see Ingham and Wiens 2021). Brennan's comparison between democracy and epistocracy could plausibly depend on which of the two tests is specified to disambiguate his model of epistocracy.

Obscure comparisons are problematic in large part because they create the risk of (perhaps only inadvertent) equivocation, in which a theorist relies on different, incompatible resolutions of the ambiguous details at different steps. We will discuss this problem in more detail below, in the context Cohen's camping trip comparison.

1.2. Mismatched comparisons A theorist makes a *mismatched comparison* when the arrangements under consideration are assumed to operate within different contexts.

Consider Simon Caney's argument for democratic global institutions, which compares two types of institutional schemes: a system of sovereign states versus a global system of suprastate institutions that "stand over and above 'states'" (2005, 161). Caney's *system of states* is basically the international status quo: a system in which states have supreme and comprehensive legal authority over a territorially defined group of people (149–50). His preferred *global system* is a framework of "democratically elected global and regional supra-state political authorities" that constrain the authority of states (161). Importantly, Caney's global system is one in which no single institution has supreme authority (163). Caney compares the two systems along two dimensions: first, the extent to which the two systems satisfy individuals' "democratic right to be able to affect the social-economic-political system in which they live and which determines what they are able to do" (156);

2 Elsewhere, Brennan proposes a voter qualification exam, which plausibly measures knowledge via exam performance (Brennan 2016, 211–12). We are unaware of any proposals to test for good faith.

second, the extent to which they advance "cosmopolitan goals" such as "the protection of human rights, the securing of a healthy environment, and the observance of cosmopolitan ideals of distributive justice" (159). He concludes that "cosmopolitan political institutions are supported on both…grounds" (163).

Caney's arguments suggest that, as he imagines them, the two systems operate in different contexts. To illustrate our point, we focus on his argument that a global system would more effectively realize cosmopolitan principles of distributive justice than a states system (159–60). Under a states system, compliance with such principles is voluntary, and there are no mechanisms for enforcing compliance. Absent enforcement mechanisms, Caney worries that "states may simply not want to secure these cosmopolitan goals" and, even if they did, "they face enormous problems posed by transaction costs, assurance dilemmas, and prisoner's dilemmas" (167, also 160). A global system overcomes these compliance and coordination problems by creating democratically elected supra-state institutions with the authority to compel compliance (161–62). But why should we expect the leaders who are charged with overseeing global institutions to use their power to compel compliance with cosmopolitan principles of distributive justice? When arguing for the need to create global enforcement mechanisms, Caney assumes that "states" — that is, the policy-makers in states — "may simply not want to secure these cosmopolitan goals." It's not clear how he could sustain his optimism about a global system were he to assume that it, too, operates in contexts where policy-makers in powerful supra-state economic institutions (or those who are tasked with holding them accountable) "may simply not want to secure…cosmopolitan goals."

Caney could reply that he is not assuming global policy-makers will be committed cosmopolitans but instead that, if placed within *democratic* supra-state institutions, electoral incentives will hold them accountable to pursue cosmopolitan goals (162). Yet, elsewhere, Caney reveals that his expectations about the efficacy of democratic global institutions assume global civil society organizations capable of fostering a robust "transnational democratic culture" (173). If we hold such contextual factors constant across the two systems, then his objections to a system of (democratic) states seem tenuous. At a minimum, Caney gives us no reason to think that a robust global civil society, capable of getting citizens and political leaders to pursue cosmopolitan goals within a global system, could not also support an effective states system by persuading citizens to hold their domestic political leaders accountable for pursuing cosmopolitan goals and thereby promoting states' voluntary compliance with cosmopolitan principles.[3]

A theorist commits the *fallacy of mismatched comparison* when they infer that one

---

3 Caney considers these possibilities but simply replies that "a statist order possesses certain deep

arrangement is better (in some respect) than an alternative on the basis of a mismatched comparison; schematically:

$$\langle A, C \rangle \text{ is better than } \langle A', C' \rangle$$
$$\therefore \quad A \text{ is better than } A'$$

The inference is invalid because the comparative judgment in the premise reflects the differences between the two models in their entirety—both their structural and their contextual features. Accordingly, the judgment in the premise may reflect not only the differences between institutional arrangements but also the differences between the contexts in which those arrangements operate. For instance, if we judge Caney's global system is better than the states system, that judgment may reflect the differences between the structural features of the institutional schemes; or it may reflect the desirability of a robust global civil society that is capable of fostering voluntary behavior in pursuit of cosmopolitan ends, which he assumes to be a feature of the context in which his global system operates but not of that in which his system of states operates.

The fallacy of mismatched comparison is loosely analogous to a violation of what empirical researchers call "internal validity," which concerns the extent to which we can can be confident that the results of an empirical study imply a causal relationship between the study's variables of interest (McDermott 2011). A study of the effect of tutoring on academic performance would lack internal validity if it inferred this effect from a comparison between students who receive tutoring and have college-educated parents and students who do not receive tutoring and have parents with no more than high school education. One cannot conclude from such a study that the better academic performance of the first group is explained (causally) by their tutoring because it fails to hold fixed a contextual factor that might also (causally) explain the difference. Analogously, one cannot conclude that a difference between institutions explains (causally or otherwise) the greater desirability of one scenario (with respect to a value like justice) compared with another unless one holds fixed the contextual factors that might also explain (causally or otherwise) its relative desirability.[4]

To show that a theorist commits the fallacy of mismatched comparison, one must

---

structural features that frustrate cosmopolitan ideals" (172). Yet, his earlier discussion of these "structural features" (166–70) neglects to consider how a robust cosmopolitan-oriented civil society might mitigate or even neutralize the anti-cosmopolitan effects of these troublesome features.

4 We note, in passing, that this is merely an illustrative analogy, not an application of the concept of internal validity to normative political theory. Political theorists do not typically make causal claims, so the standards of causal inference do not apply in any literal sense.

show that they fail to hold fixed their *primitive* (or *basic*) assumptions about contextual factors; variations in *derivative* (or *induced*) contextual factors is not enough for the charge to apply. To appreciate the point, consider Christopher Freiman's argument that Rawls "fails to make apples-to-apples institutional comparisons" when assessing the desirability of a libertarian regime, which Rawls calls "laissez-faire capitalism" or "the system of natural liberty" (Freiman 2017, 303f).[5] Freiman starts by noting that Rawls rejects laissez-faire capitalism as inconsistent with justice because it "fails to make robust economic guarantees" (Freiman 2017, 205; cf. Rawls 2001, 137). But, he argues, if we compare the two regimes while holding fixed the same idealizing behavioral assumptions, then a libertarian society need not be any worse than an liberal egalitarian society at effectively guaranteeing an adequate minimum. Freiman has us imagine a laissez-faire system in which the government solicits voluntary transfers from the more affluent, sending them instructions by email so they can "directly transfer a fair share from their paycheck to an anonymous poor recipient" (306). Although this voluntary redistributive scheme may not be a constitutive feature of laissez-faire capitalism, one might argue it would likely emerge provided "people are fully committed to distributive justice" and that "everyone supports the redistribution" (307). Such a commitment is indeed part of Rawls's description of a society that is well-ordered according to his liberal-egalitarian principles of justice. If Rawls were to hold fixed this assumption about individuals' motivations, it seems he would have to admit that any libertarian arrangement would develop a (non-coercive) institution to guarantee an adequate social minimum. His argument thus seems to presuppose an "apples-to-oranges" comparison in which liberal egalitarian arrangements are considered under idealizing assumptions about individuals' moral commitments while laissez-faire capitalist arrangements are considered under more realistic behavioral assumptions (303).

But there is only a semblance of inconsistency here. Rawls holds fixed individuals' *primitive* motivations; what varies is only their *induced* behavior, that is, the behavior that results from the assumed primitive motivations across different institutional environments. In both scenarios, people are assumed to have a "normally effective sense of justice," a broad capacity "to understand and apply the publicly recognized principles of justice, and for the most part to act accordingly as their position in society, with its duties and obligations, requires" (Rawls 2001, 9). Crucially, the specific content of this capacity depends on the institutional structure of the society an individual inhabits, just

---

5 Freiman uses three examples to press this claim against Rawls. For brevity, we focus only on the example pertaining to economic sufficiency, as our purpose is not to rebut Freiman's critique but only to register a point about what it takes to vindicate a charge of mismatched comparison.

as whether an individual's capacity for language becomes a capacity to speak English or Spanish depends on the environment in which it develops. If people inhabit a liberal egalitarian society, then they will develop and generally act in accordance with a liberal egalitarian sense of justice; if they inhabit a libertarian society, they will develop and generally act in accordance with a libertarian sense of justice (Rawls 2001, 9). Thus, Rawls does not assume what Freiman says he assumes, namely, that people are motivated to realize a specifically egalitarian distribution of goods, nor even that they are motivated to ensure that everyone has an adequate minimum. He instead holds fixed individuals' basic motivation to act in accordance with the principles of justice that are publicly recognized in the society they inhabit. Because the behavioral effects of primitive motivations are contingent on the institutional environment, it would be a mistake to hold fixed an induced behavior (such as a willingness to sacrifice income to provide others with a basic minimum) when comparing liberal-egalitarian with libertarian institutional arrangements.

1.3.   Irrelevant comparisons   A theorist makes an *irrelevant comparison* when their comparison assumes contextual factors that differ in important respects from those they "should" assume given their theoretical aims. Since questions of relevance are always a matter of disagreement among theorists, examples are most readily found in *allegations* of irrelevant comparisons rather than in the comparisons themselves.

Some of the most prominent allegations in this vein appear in debates about so-called "ideal theory." For one example, Charles Mills critiques Rawls's argument for his principles of justice for assuming a "carefully sanitized" context that excludes a history and ongoing legacy of racial injustice (Mills 2017, 37; also 75–6). If we follow Mills in accepting that "the ultimate point of ethics is to guide our actions and make ourselves better people and the world a better place," then we must reject ideal theory in favor of nonideal theory (77; also 39). Mills puts the charge this way: "What is ideally called for under ideal circumstances is not, or at least not necessarily, what is ideally called for under nonideal circumstances" (158; see also Farrelly 2007, 845). In other words, if Rawls intends to provide practical guidance in the real world, as he says he does (e.g., Rawls 1999, 216), then according to Mills, his argument for his two principles involves an irrelevant comparison between idealized well-ordered societies (i.e., between utilitarian and Rawlsian societies assuming similarly idealistic circumstances).

A theorist commits the *fallacy of irrelevant comparison* when they conclude that one arrangement is better (in some respect) than an alternative after comparing the arrangements assuming contextual factors that are ill-suited to their theoretical purpose. For example, suppose a theorist wishes to argue that one institutional arrangement *A* is a

better goal for real-world political action than an alternative A'. Given this aim, suppose (following Mills) they should compare the arrangements assuming a context C, which closely resembles real-world contexts — that is, their theoretical purpose requires them to compare $\langle A, C \rangle$ and $\langle A', C \rangle$. Suppose, however, that the theorist has instead (as Mills says of Rawls) compared the arrangements assuming a highly unrealistic context — they have concluded that A is a better goal for real-world political action than A' from a comparison of $\langle A, C' \rangle$ and $\langle A', C' \rangle$. The inference from the premise to the conclusion is invalid because an arrangement's contingent features can change depending on the context in which it operates and, insofar as our normative judgments attend to an arrangement's contingent features, changes in these features can change our comparative evaluation of arrangements.[6]

Vindicating a charge of irrelevant comparison requires carefully specifying the questions in dispute. One might be tempted to complain that a theorist's argument fails to speak to the question they intend to answer, when in fact the theorist is only guilty of not taking more interest in the questions *the critic* believes they should answer. For example, to vindicate his charge that Rawls makes an irrelevant comparison, Mills must show that Rawls intends to use his comparison of idealized well-ordered societies to support the claim that *even in non-ideal contexts*, institutions that satisfy Rawls's principles are better than alternatives. Mills attributes this intention to him (Mills 2017, 86–9, 156–60), but fails to take seriously Rawls's own characterization of his aims. In Rawls's words, he intends to identify principles that characterize "a perfectly just society," which in turn "set up an aim to guide the course of social reform" in non-ideal circumstances (Rawls 1999, 215–16).[7] This is perfectly consistent with the claim that it may not be best to satisfy Rawls's principles in "less than favorable circumstances" while we search for ways to change our circumstances on the way to realizing justice in the long term (Rawls 1999, 216; see also Gilabert 2017, 113–123; Simmons 2010, 22–23).

2. AVOIDING UNFAIR COMPARISONS

To avoid the first kind of unfair comparison, one must take care to specify clearly and consistently the important features of the models under comparison — that is, the features that affect our comparative assessment. To avoid the second, one must take care that the models under comparison all share in common those features that "should" be held

---

6   This point is often made under the rubric of "the problem of second best". See, e.g., Brennan and Pettit 2005; Coram 1996; Goodin 1995; but also Wiens 2016.

7   This is Tommie Shelby's response to Mills's critique (Shelby 2013, 150, 155–157).

fixed. And to avoid the third, one must take care to construct models that are relevant given one's theoretical aims.

Described at this level of generality, what theorists ought to do is straightforward and presumably uncontroversial. But in practice theorists routinely commit, or give their interlocutors the impression of committing, the three kinds of errors. What needs to be specified precisely to avoid the charge of an obscure comparison is not always obvious. What needs to be held fixed in a comparison is often up for reasonable debate, as are judgments about the relevance of the comparison to a theorist's larger aims.

There are no general-purpose templates or instructions for constructing good arguments in political theory. Nonetheless we think it is possible to cultivate a mindset that encourages a more structured approach to normative comparisons of institutional arrangements. We refer to what we have in mind as a *modeling mindset*, as it is exemplified by — although not equivalent to — game theorists' approach to constructing and analyzing models.

Game-theoretic models are a narrative device used to tell a kind of story (Rubinstein 2012, 19; cf. Cartwright 1991; Johnson 2021). Like any genre of story-telling, game-theoretic modeling has its own stylistic conventions and guidelines. A game theorist must respect the rules of game-theoretic story-telling, just as a poet composing iambic poetry must adhere to the genre's constitutive rules of composition. In particular, a well-defined game-theoretic model must clearly specify:

- the set of relevant actors in the story ("players");

- the plans of action available to them ("strategies");

- their initial beliefs about their situation, including the possible consequences of their actions;

- their primitive preferences, which encode their basic goals, ethical commitments, or other evaluative attitudes.

These are the basic ingredients from which the story-teller develops explanations of the characters' behavior. They are the *exogenous* elements of the model, in the sense that the analyst stipulates them and does not try to explain them by reference to other elements of the model. What the characters will do and what will result from their behavior are *endogenous*, in the sense of being explained by reference to the model's exogenous elements. The rules for explaining the characters' behavior in terms of these ingredients — the rules for inferring which forms of behavior and which outcomes are

consistent with the exogenous elements — are encoded in "solution concepts," such as Nash equilibrium or subgame perfect equilibrium.

The qualifier before "preferences" is important. An actor may have a *primitive* preference for, say, competence in their elected officials, which only becomes an *induced* or *derivative* preference for one party's candidate over the other after the actor has acquired information and updated her beliefs about the two parties' candidates. The primitive preference, which is exogenously given, serves, together with the other features of the model, to explain the endogenously induced preference. Nothing in our characterization of models, or in the usual rules for explaining behavior in game-theoretic models, prevents the analyst from telling a story in which the characters' beliefs and preferences change as the story unfolds. But to ensure that the story unfolds in a manner that is consistent with the genre's rules for explaining the characters' behavior, the analyst must carefully distinguish the model's exogenous features from its endogenous features and be mindful of the need to justify their claims about the latter by reference to assumptions about the former.

The last sentence indicates some of game theorists' intellectual dispositions, which exemplify the more general modeling mindset we claim is useful for avoiding unfair comparisons: a disposition to conceptualize social situations in terms of certain basic ingredients; a disposition to distinguish the model's exogenous elements from its endogenous elements; and a disposition to justify claims about the model's endogenous elements by reference to its exogenous elements in accordance with the rules that govern behavior within the model world. These dispositions readily generalize beyond the use of game-theoretic models in particular to the construction and analysis of any type of model, including the kinds of informal models normative theorists often compare for the purpose of explicating our reasons to prefer one type of social and political arrangement over alternatives. By a *modeling mindset*, we just mean these general dispositions, of which the game-theoretic variety is a special case.

3.   THE MODELING MINDSET: AN ILLUSTRATION

To illustrate how a modeling mindset enables one to avoid the unfair comparisons outlined above, we develop an extended example using G.A. Cohen's (2009) comparison of two models of a camping trip. Cohen's thought experiment is useful because critics have charged him with committing each of the three fallacies associated with unfair comparisons. We present a version of his thought experiment that clears him on all counts, while arguably still illustrating certain respects in which socialism is more desirable than capitalism, provided it is feasible. To be clear, we do not intend to defend

Cohen's conclusion about the relative merits of socialism. We aim to show only that he could have avoided various charges of making an "unfair" comparison had he taken a more structured approach to his thought experiment.

In *Why Not Socialism?*, Cohen argues for two socialist principles of justice using a comparison between two versions of a camping trip: a "socialist" camping trip and a "capitalist" trip. The trips differ in the manner in which the campers allocate their resources. In the "socialist" campground, the campers place whatever resources they have (including time and talents) in a common stock under collective control, and each makes use of these collective resources as they need (3–4). In the "capitalist" campground, the campers retain private property rights over their personal resources, and the use of these resources is subsequently determined on the basis of bargaining and exchange (5–6). Cohen's purpose is to elicit in his readers the intuition that a camping trip run on socialist principles is more desirable than one run on capitalist principles and then infer from this intuition that a socialist society is more desirable than a capitalist society (10, 46–52).

Commentators have charged Cohen with an *obscure* comparison. As Sharon Krause puts it:

> [The] failure [of Cohen's argument] results in part from a conceptual ambiguity that runs throughout the book with respect to the meaning of socialism. Sometimes socialism is defined in terms of the specific economic practice of collective ownership; at other times, Cohen equates it with the more general moral principles of equality and community, and with the dispositions of generosity, friendship, cooperation, and unselfishness. Cohen's argument for the moral desirability of socialism trades on this ambiguity. As in the camping trip example, the moral appeal of equality and community, generosity and friendship, is taken to justify the economic system of collective ownership. (Krause 2010, p. 886)

To avoid Krause's charge, Cohen must answer questions about the basic features of the two scenarios, such as:

- Is he stipulating that the "socialist" campers have a *primitive* preference to share what they produce and to treat their productive capacities as collective resources for the benefit of the group? Or is their willingness to share an *induced* preference — explicable in terms of, say, more primitive preferences to enjoy the benefits of social cooperation together with beliefs about what other campers

14

expect of them as a condition of social cooperation?[8]

- When imagining a capitalist version of the camping trip, is Cohen asking us to entertain alternative assumptions about the campers' primitive preferences? Or is he asking us to imagine campers with the same primitive preferences as in the socialist version but interacting within an alternative set of social norms, which induce the behaviors he associates with capitalism?

Jason Brennan (2014) takes advantage of the ambiguity in Cohen's comparison to charge him with the fallacy of *mismatched* comparison. On Brennan's interpretation, Cohen assumes that the socialist campers are fundamentally disposed to contribute their time and energy to the collective resource stock and to allow others to draw on those resources as they need, whereas he assumes that the capitalist campers are fundamentally disposed to claim the products of their labor as private property and bargain with other campers in pursuit of personal gain. As Brennan puts it, "Cohen is not comparing like to like" because he compares an "idealized" socialist camping trip "with morally perfect people" who are (by stipulation) "motivated by fellow-feeling, community spirit, and love" to a "non-idealized, realistic" capitalist camping trip with "real, flawed people" who are (by stipulation) motivated by "selfish[ness] and fear[]" (Brennan 2014, 58–59).

Even if Cohen can address these two concerns, critics have raised a third: that a comparison of socialist and capitalist camping trips is *irrelevant* for the purpose of assessing socialism and capitalism as large-scale social systems. Cohen anticipates this objection (2009, 10, 49–50, 53f), but critics argue that his response falls short. For example, Miriam Ronzoni points to the fact that participating in the camping trip is voluntary whereas participating in mass society is not: unlike participants in mass society, campers opt into an activity for a limited duration, the express purpose of which is to realize the values of equality and community (2011, 172, 175). David Schmidtz points to differences in resource availability (resources are more abundant relative to people's desires on camping trips than in modern societies) and personal familiarity of the kind that sustains trust among participants (2011, 787). "[W]hat we need to imagine," according to Schmidtz, "is not friends on a camping trip but a campground where strangers come and go" (ibid.).

These critiques raise a challenge: Can we present a thought experiment that preserves the spirit of Cohen's original while avoiding the charges of obscure, mismatched, and irrelevant comparisons? Before trying to meet this challenge, we set out some criteria for success.

---

8   We interpret Cohen's remarks on pp. 4f as suggesting something closer to the latter.

To remain faithful to Cohen's purpose, our model must satisfy at least three conditions:

(i) It should describe a situation in which two or more individuals can cooperate for mutual gain but face choices about the terms on which to cooperate.

(ii) One of their options is to cooperate on terms that are relevantly analogous to "socialism," and a second is to cooperate on terms that are relevantly analogous to "capitalism."

(iii) Individuals differ in their productive capacities as well as their preferences and tastes.

To avoid the objection that socialism is ambiguously defined, we want a model in which

(iv) there is a clear conceptual distinction between the choices individuals make, the values and goals underlying those choices, and the institutions constraining those choices.

To avoid the objection that Cohen's argument relies on a mismatched comparison, the model should allow us to

(v) compare capitalism and socialism while holding fixed the relevant contextual factors, in particular the primitive assumptions about individuals' values and goals.

Finally, to avoid the objection of irrelevance, we want a model that

(vi) generalizes to a wide range of social interactions, can readily accommodate salient complications, and makes only minimal ethical assumptions, which are also neutral with respect to the comparison between socialism and capitalism.

3.1. The camping trip redux Two people embark on a camping trip. Each day, the campers collect resources and then allocate what they collect using one of two mechanisms. According to the *socialist mechanism*, they pool the resources they collect and each draws upon them as they need; for simplicity, we assume they have equal needs, which implies they enjoy equal shares.[9] According to the *capitalist mechanism*, they

---

9 An alternative socialist mechanism grants the campers private control over equal shares of resources and allows them to trade in an effort to resolve any Pareto inefficiency from a strictly egalitarian distribution. This seems consistent with, and perhaps implied by, Cohen's conception of socialist equality (2009, 18–22). Since excluding this complication makes no difference to the substance of our conclusions, we set it aside to keep things simple.

each hold the resources they collect as private possessions and trade them at a rate that equates supply (as determined by their collecting abilities) with demand (as determined by their respective tastes and preferences). If the campers cannot agree on an allocation mechanism, then no exchange takes place and each camper consumes whatever they collect. The campers differ in their tastes and preferences over bundles of goods as well as their abilities to collect goods. Finally, we assume that they comply with minimal ethical constraints: neither is willing to use force or fraud to acquire what the other has produced, thus each implicitly recognizes the other's voluntary cooperation as a necessary condition for any exchange to take place.

Here is a concrete example of the model.[10] One camper, *A*, catches three pounds of fish each day; the other, *B*, collects ten pounds of firewood. By definition, under the socialist mechanism, the campers consume equal shares of the goods collected each day (assuming equal needs). Under the capitalist mechanism, the campers get whatever shares result from trading the goods they each collect at a rate that equates supply with demand. To determine these shares, we need to make some assumptions about the campers' preferences over possible bundles of resources. As a general matter, we assume that both campers rank any mixed bundle of goods above any bundle with only one good. For the sake of concreteness, we assume that the campers' preferences are such that, under capitalism (and given their gathering abilities), *A* and *B* get equal amounts of firewood but *A* gets more fish than *B*.[11]

Our assumptions about their preferences and abilities imply that *A* prefers her capitalist bundle to her socialist bundle, while *B* prefers his socialist bundle to his capitalist bundle. These induced preferences over allocation mechanisms follow from three observations: first, that the campers receive equal amounts of firewood under both capitalism and socialism; second, that *A* receives more fish under capitalism than under socialism; third, that *B* receives more fish under socialism than under capitalism. It follows from our assumption that both campers prefer any mixed bundle to any non-mixed bundle that they both prefer the capitalist and socialist scenarios to the scenario in which no exchange takes place.

To this point, we have stipulated four types of primitive features: the available al-

---

10   This is one of many ways to fill in the details. Our choice of example makes no qualitative difference to the lessons we draw from the model.

11   We present a formal example in Appendix 1 that illustrates these informal assumptions. Other assumptions about their preferences would work equally well for the purposes of the story, provided the assumptions imply that *A* prefers her allocation under capitalism to the equal division, and *B* prefers the equal allocation to his allocation under capitalism, and each camper prefers both of these allocations to the status quo without exchange.

location mechanisms; the campers' collecting abilities; their preferences over bundles of goods; and, implicitly, their shared commitment to voluntary exchange. These assumptions resolve worries about the ambiguities in Cohen's thought experiment and, in particular, the ambiguity identified by Krause. In our model, "socialism" and "capitalism" refer to distinct mechanisms for allocating goods between the campers rather than the campers' motivational dispositions or their ethical commitments. Specifically, we do not assume that the campers are disposed to abide by "socialist norms" or "capitalist norms." Indeed, with respect to their primitive motivations, we make only two assumptions: that the campers care about their own consumption, and that they are committed to ensuring that any exchange of goods is based on voluntary compliance with an agreed-upon allocation mechanism. Importantly, neither camper has, by stipulation, an ethical commitment to socialism, nor to any other-regarding behaviors that might be associated with socialism. While it is true that B prefers the outcome of the socialist mechanism and thus has an instrumental reason to prefer its use, his preference is *induced* rather than primitive — it is contingent upon and explained by appeal to more fundamental features of the model. To wit, we could make B prefer the capitalism mechanism instead by simply varying our assumptions about the players' productive abilities and tastes while holding all else fixed. Neither camper's preference over allocation mechanisms is "moral" in any sense. If the campers behave as "socialists" or "capitalists," this behavior will emerge as a consequence of their efforts to satisfy their consumption preferences through voluntary exchange. Socialist and capitalist behavior are thus endogenous rather than exogenous features of the model.

These observations go part of the way toward addressing Brennan's critique: if there turns out to be anything desirable about a scenario in which the campers use the socialist mechanism, it will not be in virtue of the fact that we, the analysts, have populated the scenario with morally idealized agents. To go the rest of the way, we must show that the assumptions about the campers' primitive motivations are compatible with both socialism and capitalism, and that the model permits a comparison of the two institutions while holding fixed the campers' primitive motivations. To do this, we show that it can be instrumentally rational for our campers to agree to use either the socialist mechanism or the capitalist mechanism, depending on their beliefs about the other camper's intentions. Not only does this explanation refrain from appealing to shared ethical commitments — as we have emphasized, we do not assume the campers have any ethical commitments beyond the minimal one of treating each other's voluntary agreement as necessary for exchange — it is also compatible with supposing, as we do, that the campers have conflicting preferences over the mechanisms, rooted in their

self-regarding consumption preferences and unequal productive capacities.

Recall that the campers must agree on an allocation mechanism for any exchange to take place. We assume a simple procedure for settling on a mechanism. Each day, after collecting their goods, the campers meet and *A* proposes a mechanism, which *B* can accept or reject; if *B* accepts the proposal, then the campers exchange goods as dictated by the agreed upon mechanism; if *B* rejects *A*'s proposal, then no exchange takes place and each camper consumes whatever they collected that day. The "socialist" camping trip corresponds to a scenario in which the campers consistently agree to the socialist allocation; the "capitalist" camping trip corresponds to a scenario in which the campers consistently agree to the capitalist allocation.[12]

Both camping trips can result from rational behavior in the sense that each is compatible with both campers implementing a plan of action that delivers their most preferred allocation given the other camper's plan of action.[13] In the socialist version, *A* proposes the socialist mechanism each day, unless *B* accepted the capitalist mechanism on the previous day, in which case *A* proposes the capitalist mechanism again; *B* always accepts a proposal of the socialist mechanism, and accepts a proposal of capitalism if and only if he accepted a proposal of capitalism on the previous day (which never happens, in equilibrium). In the capitalist version, *A* always proposes the capitalist mechanism, while *B* accepts either proposal from *A*. Crucially, our assumptions about the campers' fundamental motivations and preferences are the same in either scenario, so there is now no basis for Brennan's charge of a mismatched comparison.[14]

Having explained why our version of the thought experiment preempts the charges of obscure and mismatched comparisons, let us now consider how it handles the charge of irrelevance. Cohen offered the camping trip as an example of a context in which, intuitively speaking, a socialist mode of organizing production and allocation is desirable and in which this intuitive response is explained by socialism's realization of two principles, one concerning community and one concerning equality of opportunity. He then claims that the intuitive appeal of realizing those principles on the camping trip is not contingent on any circumstances specific to the camping trip, and that it would be desirable to realize them on a national scale provided it were feasible. However, his critics have questioned the relevance of his thought experiment on the grounds that it neglects

---

12 All our conclusions hold if we instead assume that *B* proposes a mechanism, which *A* accepts or rejects.

13 Technically, both versions of the camping trip can be shown to be not only Nash equlibria but also *subgame perfect equilibria*. See Appendix 3.

14 Some might doubt whether the claims in this paragraph are consistent with the assumptions about the campers' primitive preferences and abilities made earlier in the text. To address such doubts we present a formal example in Appendix 3.

facts about large-scale social institutions, in particular the non-optional character of participation in mass society, resource scarcity, and lack of personal familiarity and trust.

Our reconstruction of Cohen's thought experiment can help us reason clearly about the question of relevance. Within our model, the socialist mechanism realizes Cohen's principle of equal opportunity: the morally arbitrary, unearned differences in native productive abilities do not produce inequalities in consumption. The capitalist mechanism does not: A ends up consuming more fish, simply in virtue of the difference in the campers' natural talents. Our reconstruction allows us to see that *this* conclusion about the model does not rest on its interpretation as a model of a camping trip. It merely registers a simple, general truth about market allocations: differences in productive abilities, even when they reflect morally arbitrary differences, produce differences in consumption under standard assumptions about the market mechanism. If one has the intuition that the socialist scenario is more desirable than the capitalist in virtue of this difference when the model is interpreted as a camping trip, we suspect the intuition will survive when the model is reinterpreted in some other way — interpreting A and B as, say, two shipwrecked island castaways instead. Keep in mind that under any interpretation of the model, the socialist and capitalist outcomes arise as a result of voluntary, instrumentally rational behavior. The intuition we (on Cohen's behalf) mean to invoke is not that it would be desirable for B to compel A to cooperate on socialist terms with threats of violence. It is rather the intuition that it would be desirable if individuals voluntarily agreed to use a mechanism that does not convert morally arbitrary differences in productive abilities into inequalities in consumption.

If that intuition survives even when the the two individuals are reinterpreted as ship-wrecked castaways, then one cannot object to the relevance of the comparison between the socialist and capitalist camping trips on the grounds that campers are typically acquaintances, while the members of a large society are strangers. Our reconstruction of Cohen's thought experiment also abstracts from other features of camping trips that lead critics to question its relevance, thereby preempting their objections. Within our model the campers cannot genuinely choose whether to participate in repeated interactions with each other. They have no "outside option" other than the option of refusing to engage in any exchange using either allocation mechanism, which is prohibitively costly. We also assume resources are scarce: the campers have unlimited appetites but are capable of producing only a finite quantity of goods to satisfy them. Thus we can be confident that any conclusions about the socialist mode of organizing "the camping trip" are not driven by implicit assumptions of voluntary participation or resource abundance.

Our model scenario does, however, abstract from facts about socialism and capi-

talism that are surely relevant to an all-things-considered judgment about their relative desirability in large-scale settings. For example, it assumes that production is costless and that the only problem confronting "society" is which institutional mechanism to use for allocating the goods that individuals produce at no cost. A familiar concern with socialism is that it fails to incentivize efficient productive activity because it does not sufficiently reward costly exertions of effort. Thus a next step would be to consider a variation on our scenarios in which the campers can produce goods only at some cost. If the socialist mechanism then fails to realize the principle of equal opportunity more effectively than the capitalist mechanism, or if the revised thought experiment no longer generates the intuition that socialism is more desirable, then one could conclude that the rhetorical effect of the original thought experiment relies on abstracting from this consideration, undermining its relevance to the broader comparison between socialism and capitalism. Alternatively, one might find that the intuition survives this modification to the model, in which case the observation about production costs fails to vindicate a charge of irrelevance.

One virtue of the modeling mindset is that it leads one to impose a modular conceptual structure on thought experiments, which facilitates this kind of inquiry. Our version of Cohen's thought experiment has an abstract conceptual structure — two actors, who at no cost produce goods of value to both, and who can jointly choose one of two mechanisms for allocating the goods between them, etc. — and an informal narrative component — a story about two campers — laid over it. One can check whether one's intuitions reflect circumstances specific to camping trips just by swapping out the camping trip story for an alternative interpretation of the more abstractly described model, knowing that even after relabeling and reinterpreting the identities of the actors, the underlying structure of their situation will remain unchanged. There will still be two possible outcomes consistent with the operative assumptions about the actors' rationality and motivations, and it will still be true of only one of these that differences in productive capacities are not converted into inequalities in consumption. By contrast, if one wishes to check the sensitivity of conclusions to primitive behavioral assumptions or other assumptions that define the abstract structure of the model — such as the assumption that it is costless to produce goods — one must analyze a variant of the model that results from modifying these assumptions.

4. CONCLUSION

Political theorists aspire to answer comparative questions — which institutional arrangements best realize the value of equality? which social structures best satisfy principles of

21

justice? — and often use comparisons between hypothetical scenarios to this end. We have noted three ways in which inferences from these comparisons can go awry — comparisons may be obscure, mismatched, or irrelevant — and explained how a modeling mindset can help one guard against these errors.

What we are recommending is a mindset, not a foolproof procedure or method that can be mechanically applied. That is, we are recommending a particular set of dispositions rather than a recipe or algorithm — dispositions to conceptualize social situations in terms of clearly defined primitive ("exogenous") elements, to distinguish between these and the ("endogenous") behaviors and outcomes they explain, and to justify claims about the latter in terms of assumptions about the former. Our reconstruction of Cohen's thought experiment illustrates both the promise of this mindset and its limitations. The promise is that these dispositions bring in their wake a disposition to think more explicitly and self-consciously about how to define critical concepts (mitigating the risk of obscure comparisons); what should be held fixed and what should be allowed to vary across the alternatives (mitigating the risk of mismatched comparisons); and whether the comparisons within the thought experiment, despite its deliberate abstraction from and simplification of real-world complexities, are relevant to theoretical questions of interest (mitigating the risk of irrelevant comparisons). In all cases, however, the promise is just that a modeling mindset encourages us to raise certain questions; the answers to these questions would then be up for debate. A modeling mindset cannot ensure we construct thought experiments that withstand critical scrutiny; we still need imagination and judgment. But a modeling mindset can direct our imagination and judgment in ways that help us avoid recurrent reasoning errors.

5. APPENDIX

5.1. The capitalist example in detail.   We derive each camper's resource bundle under capitalism from primitive assumptions about their preferences over resource bundles and their resource collecting abilities. We represent their preferences over possible bundles of goods using the following utility functions:

$$u_A(x_A, y_A) = x_A^{0.75} y_A^{0.25} \qquad\qquad u_B(x_B, y_B) = x_B^{0.5} y_B^{0.5},$$

where x_A and x_B denote each camper's share of fish and y_A and y_B denote their shares of firewood. These utility functions are simply accounting devices; they encode each camper's comparative evaluation of possible bundles of goods. The particular functions we have chosen are instances of the Cobb-Douglas utility function, which is commonly

used in consumer theory; in general it has the form $u(x, y) = x^\alpha y^{1-\alpha}$, where the parameter α is a number between 0 and 1 (see Mas-Colell, Whinston, and Green 1995, ch. 3). Note that the functions imply the assumption we make in the main text: each consumer prefers any mixed bundle to any bundle with only one of the goods because $u_i(x, y) > 0 = u_i(0, z) = u_i(z, 0)$ for any $x, y, z > 0$.

As for their collecting abilities, recall our assumption that, each day, A collects three pounds of fish and B gathers ten pounds of firewood.

From these primitive assumptions about preferences and gathering abilities, we can calculate the "market price" and, in turn, the campers' bundles under capitalism in several steps. First, for each camper, we take the derivates of their utility function with respect to x and y and subject to the constraint that the overall cost of the bundle they demand does not exceed their "budget," conceptualized as the market value of the goods they collect prior to trading. Precisely, we solve the following maximization problem for $i \in \{A, B\}$, prices p_x and p_y and initial endowments $\hat{x}_i$ and $\hat{y}_i$ (i.e., amount of each good collected by i prior to trading):

$$\max_{x_i, y_i} u_i(x_i, y_i)$$

subject to the budget constraint $p_x x_i + p_y y_i \leq w_i$, where w_i is the "market value" of the goods i is able to gather on their own, i.e., $w_i = p_x \hat{x}_i + p_y \hat{y}_i$. To reduce notation, we can normalize the price of fish to $p_x = 1$ without affecting any of our conclusions. Player A's budget constraint is then $x_A + p_y y_A \leq 3$, and player B's $x_B + p_y y_B \leq 10 p_y$. Solving the optimization problems gives us the campers' demand functions for each good (for the derivation see Mas-Colell, Whinston, and Green 1995, pp. 55, 56):

$$x_A^*(p_x, p_y, w_A) = \frac{3w_A}{4p_x} = \frac{9}{4} \qquad\qquad x_B^*(p_x, p_y, w_B) = \frac{w_B}{2p_x} = 5p_y$$

$$y_A^*(p_x, p_y, w_A) = \frac{w_A}{4p_y} = \frac{3}{4p_y} \qquad\qquad y_B^*(p_x, p_y, w_B) = \frac{w_B}{2p_y} = 5$$

The next step is to set the sum of the two campers' demand functions for a good (say, y) equal to the total supply of that good (which is 10 in the case of y) and solve for the prices that balance supply and demand:

$$y_A^*(p_x, p_y, w_A) + y_B^*(p_x, p_y, w_B) = 10 \implies \frac{3}{4p_y} + 5 = 10 \implies p_y = \frac{3}{20}$$

With this price ratio in hand, we calculate the campers' post-exchange bundles by plugging the prices into their demand functions:

$$x_A^* = \frac{9}{4} \qquad\qquad\qquad x_B^* = \frac{3}{4}$$

$$y_A^* = 5 \qquad\qquad\qquad y_B^* = 5$$

5.2. Campers' mechanism preferences. Let (x, y) denote a bundle of goods, with the first number indicating the amount of fish and the second number indicating the amount of firewood. Under socialism, both campers get $(1.5, 5)$; under capitalism, A gets $(2.25, 5)$ and B gets $(0.75, 5)$. Absent exchange, A gets $(3, 0)$ and B gets $(0, 10)$. By plugging these allocations into the campers' utility functions, the reader can confirm

$$u_A(2.25, 5) > u_A(1.5, 5) > u_A(3, 0) \qquad u_B(1.5, 5) > u_B(0.75, 5) > u_B(0, 10)$$

5.3. Camping equilibria. We show that the two camping trips described in the text are *subgame perfect equilibria*. For $i \in \{A, B\}$, let u_i^c and u_i^s denote camper i's payoff from the capitalist and socialist allocations respectively. Following the example in Appendix 5.2, we assume that $u_A^c > u_A^s > 0$ and $u_B^s > u_B^c > 0$.

We model the camping trip as an indefinitely repeated game: each day, A proposes an allocation mechanism — either "socialism" or "capitalism" — which B accepts or rejects. If B accepts A's proposal, then they exchange goods as dictated by the agreed upon mechanism; if B rejects A's proposal, then no exchange occurs. The same process repeats the next day. We let $\delta \in (0, 1)$ represent the campers' shared concern for the value of future interactions, with campers caring relatively more about the future as δ gets closer to 1.

Socialist version: A proposes socialism everyday unless B accepted a proposal of capitalism on the previous day, in which case she proposes capitalism again; B always accepts a proposal of socialism, and he accepts a proposal of capitalism if and only if he accepted a proposal of capitalism on the previous day.

These strategies form a subgame perfect equilibrium provided B is sufficiently patient. To see this, observe that when the players follow these strategies, B's stream of payoffs is $u_B^s + \delta u_B^s + \cdots = \frac{u_B^s}{1-\delta}$. Clearly B has no incentive to deviate by rejecting any proposal of socialism: deviating in this way and then reverting to the postulated strategy, his stream of payoffs (from the deviation onward) would be $0 + \delta u_B^s + \cdots = \frac{\delta u_B^s}{1-\delta} < \frac{u_B^s}{1-\delta}$. Off the path of play, B also has no incentive to accept any offer of capitalism (and then revert to the postulated

strategy), provided he is sufficiently patient: for then his stream of payoffs (from the deviation onward) would be $u_B^c + \delta u_B^c + \cdots = \frac{u_B^c}{1-\delta}$, whereas by following the postulated strategy his stream of payoffs is $0 + \delta u_B^s + \cdots = \frac{\delta u_B^s}{1-\delta}$. Thus, provided $\delta \geq \frac{u_B^c}{u_B^s}$, player B cannot profitably deviate from his strategy in any subgame, by the one-stage deviation principle.

Now turn to player A. On the path of play her stream of payoffs is $u_A^s + \delta u_A^s + \cdots = \frac{u_A^s}{1-\delta}$. If she were to deviate at any decision node by proposing capitalism and then revert to the postulated strategy, the offer of capitalism would be rejected and her subsequent offers of socialism accepted, and her stream of payoffs would be $0 + \delta u_A^s + \cdots = \frac{\delta u_A^s}{1-\delta} < \frac{u_A^s}{1-\delta}$. Turning to subgames off the path of play, consider any subgame arising after B has accepted a proposal of capitalism in the previous period. If A follows her postulated strategy by proposing capitalism, then her stream of payoffs going forward is $u_A^c + \delta u_A^c + \cdots = \frac{u_A^c}{1-\delta}$ (player B, having accepted capitalism in the previous period, continues to accept it according to his postulated strategy). If she deviates by proposing socialism instead and then reverts to her postulated strategy, her stream of payoffs would be $u_A^s + \delta u_A^s + \cdots = \frac{u_A^s}{1-\delta}$, which is worse because $u_A^s < u_A^c$. Thus, by the one-stage deviation principle, she cannot profitably deviate in any subgame.

Capitalist version: Each day, A proposes capitalism and B accepts any proposal.

These strategies also form a subgame perfect equilibrium (for any $\delta \in (0,1)$). To see this, observe that starting at any decision node at which A makes a proposal, her stream of payoffs from following the postulated strategy is $u_A^c + \delta u_A^c + \cdots = \frac{u_A^c}{1-\delta}$. If she were to deviate and then revert to the postulated strategy, her stream of payoffs would be $u_A^s + \delta u_A^c + \cdots = u_A^s + \frac{\delta u_A^c}{1-\delta}$, which is worse. If B ever deviates from the postulated strategy by rejecting a capitalist proposal but then reverting to the postulated strategy, his stream of payoffs is $0 + \delta u_B^c + \cdots = 0 + \frac{\delta u_B^c}{1-\delta}$, whereas sticking to the postulated strategy gives him $u_B^c + \delta u_B^c + \cdots = \frac{u_B^c}{1-\delta}$. Clearly no such deviation is proftable. At subgames (off the path of play) where A has proposed socialism, B's stream of payoffs from following his strategy is $u_B^s + \delta u_B^c + \delta^2 u_B^c + \cdots = u_B^s + \frac{\delta u_B^c}{1-\delta}$, while his stream of payoffs from deviating (by rejecting the offer) and reverting to the postulated strategy is $0 + \delta u_B^c + \cdots = \frac{u_B^c}{1-\delta}$, which is worse. Thus neither player can profitably deviate in any subgame.

REFERENCES

Bagg, Samuel. 2018. "The Power of the Multitude: Answering Epistemic Challenges to Democracy." *American Political Science Review* 112: 891–904.

Brennan, Jason. 2014. *Why Not Capitalism?* New York: Routledge.

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

Brennan, Jason. 2018. "Does the Demographic Objection to Epistocracy Succeed?" *Res Publica* 24: 53–71.

Caney, Simon. 2005. *Justice Beyond Borders* New York: Oxford University Press

Cartwright, Nancy. 1991. "Fables and Models." *Proceedings of the Aristotelian Society* 65: 55–68.

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

Collier, David and Gerardo Munck. 2022. *Critical Junctures and Historical Legacies: Insights and Methods for Comparative Social Science.* Lanham, MD: Rowman & Littlefield.

Coram, Bruce Talbot. 1996. "Second Best Theories and the Implications for Institutional Design." In *The Theory of Institutional Design*, edited by Robert E. Goodin. Cambridge: Cambridge University Press, pp. 90–102.

Dowding, Keith. 2022. "How to Use Imaginary Cases in Normative Theory." *Metaphilosophy* 4: 512–525.

Egami, Naoki and Erin Hartman. 2023. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review* 117: 1070–1088.

Farrelly, Colin. 2007. "Justice in Ideal Theory: A Refutation." *Political Studies* 55: 844–864.

Freiman, Christopher. 2017. "Ideal Theory." In *The Routledge Handbook of Libertarianism*, edited by Jason Brennan, David Schmidtz, and Bas Van der Vossen. New York:

Routledge.

Gilabert, Pablo. 2017. "Justice and Feasibility: A Dynamic Approach." In *Political Utopias*, edited by Kevin Vallier and Michael Weber. New York: Oxford University Press.

Goodin, Robert E. 1995. "Political Ideals and Political Practice." *British Journal of Political Science* 25: 37–56.

Ingham, Sean and David Wiens. 2021. "Demographic Objections to Epistocracy: A Generalization." *Philosophy and Public Affairs* 49:323–349.

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

Johnson, James. 2021. "Models-as-Fables: An Alternative to the Standard Rationale for Using Formal Models in Political Science." *Perspectives on Politics* 19: 874–889.

Krause, Sharon. 2010. "Beyond Capitalism?" *Political Theory* 38: 884–890.

Landa, Dimitri and Ryan Pevnick. 2021. "Is Random Selection a Cure for the Ills of Electoral Democracy?" *Journal of Political Philosophy* 29: 46–72.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford University Press.

McDermott, Rose. 2011. "Internal and External Validity." In *Cambridge Handbook of Experimental Political Science*, edited by James N. Druckman, Donald P. Green, James H. Kurlinski, Arthur Lupia. New York: Cambridge University Press.

Mills, Charles W. 2017. *Black Rights/White Wrongs*. New York: Oxford University Press.

Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. New York: Cambridge University Press.

Rawls, John. 1999. *A Theory of Justice*. Revised ed. Cambridge, MA: Belknap Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement.* Cambridge, MA: Belknap Press.

Ronzoni, Miriam. 2011. "Life is Not a Camping Trip — On the Desirability of Cohenite Socialism." *Politics, Philosophy, and Economics* 11: 171–185.

Rubinstein, Ariel. 2012. *Economic Fables.* Cambridge: Open Book Publishers.

Sartori, Giovanni. 1991. "Comparing and Miscomparing." *Journal of Theoretical Politics* 3: 243–257.

Schmidtz, David. 2011. "Nonideal Theory: What It Is and What It Needs to Be." *Ethics* 121: 772–796.

Shelby, Tommie. 2013. "Racial Realities and Corrective Justice: A Reply to Charles Mills." *Critical Philosophy of Race* 1: 145–162.

Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy and Public Affairs* 38: 5–36.

Wiens, David. 2016. "Assessing Ideal Theories: Lessons from the Theory of Second Best." *Politics, Philosophy, and Economics* 15: 132–149.

Wolin, Sheldon. 2004. *Politics and Vision: Continuity and Innovation in Western Political Thought.* Expanded ed. Princeton and Oxford: Princeton University Press.

Valentini, Laura. 2011. *Justice in a Globalized World.* New York: Oxford University Press.

Ypi, Lea. 2010. "On the Confusion Between Ideal and Non-ideal in Recent Debates on Global Justice." *Political Studies* 58: 536–555.