Version 2.1 // April 2018

The General Theory of Second Best Is More General Than You Think

David Wiens

Abstract. Lipsey and Lancaster's "general theory of second best" is widely thought to have significant implications for applied theorizing about the institutions and policies that most effectively implement abstract normative principles. It is also widely thought to have little significance for theorizing about which abstract normative principles we ought to implement. Contrary to this conventional wisdom, I show how the second best theorem can be extended to myriad domains beyond applied normative theorizing, and in particular to more abstract theorizing about the normative principles we should aim to implement. I start by separating the mathematical model used to prove the second best theorem from its familiar economic interpretation. I then develop an alternative normative-theoretic interpretation of the model, which yields a novel second best theorem for idealistic normative theory. My method for developing this interpretation provides a template for developing additional interpretations that can extend the reach of the second best theorem beyond normative theoretical domains. I also show how, within any domain, the implications of the second best theorem are more specific than is typically thought. I conclude with some brief remarks on the value of mathematical models for conceptual exploration.

There is growing recognition among political philosophers that the "general theory of second best" (Lipsey and Lancaster, 1956) has significant implications for applied normative theorizing about institutional design and policy choice across a range of social circumstances. The basic idea has become familiar enough: among options that fall short of the ideal, the best institutional scheme or policy regime does not necessarily resemble, and may radically differ from, the ideal (e.g., Brennan and Pettit, 2005; Coram, 1996; Goodin, 1995; Räikkä, 2000; Wiens, 2016). The second best doesn't necessarily look much like the ideal. What's more, it doesn't necessarily look like the best feasible approximation of the ideal.

The theory of second best is also widely thought to have little significance for normative theory beyond questions about institutional design and policy choice. In particular,

Acknowledgements. Earlier versions of this paper have been presented at Ponticifia Universidad Católica de Chile and the PPE Society annual conference; I'm grateful to audience members for helpful discussion. Thanks in particular to Dave Estlund, Sean Ingham, Cristian Pérez Muñoz, Ross Mittiga, and Julia Staffel for productive discussion.

it is thought — even by those who press its significance for applied theorizing — to be largely irrelevant for theorizing about which normative *principles* justifiably hold under various circumstances (Goodin, 1995; Räikkä, 2000; Swift, 2008). While the theory of second best impinges on our practical efforts to implement general normative principles, it leaves untouched the principles we should aim to implement. Indeed, it might be hard to see how things could be otherwise. In its native scholarly context, the theory of second best warns against the piecemeal (i.e., sector-by-sector) pursuit of Pareto efficiency as a means to maximizing social welfare: if one sector violates Pareto efficiency, realizing a social welfare maximum will typically require departing from Pareto efficiency in all other sectors. The challenge thus targets the wisdom of implementing certain policies as means to realizing a fixed normative goal, namely, maximizing social welfare. Extrapolating to broader normative contexts, the sensible lesson seems to be that the theory of second best warns us against using idealistic institutions and policies to realize our normative goals in nonideal contexts, while leaving the specification of these more general goals untouched.

I show that the theory of second best is more general than is conventionally thought. As economists recognized from the start, the original theorem is not about economic theory and welfare policy *per se* but is instead about mathematical optimization in general.[1] Yet the significance of this point is widely unappreciated. Once we distinguish between the mathematical model and associated theorem on the one hand and the familiar economic interpretation of the model on the other, we clear the way for developing novel applications of the theorem. For every plausible interpretation of the model, we get a new second best theorem, thereby extending the reach of the theory of second best beyond the domain of applied social (including economic and political) theory.[2]

Having opened the door to new interpretations, I extend the application of the theory of second best beyond applied normative reasoning to more abstract forms of normative reasoning. To do this, I develop a model of normative reasoning about ideal principles, i.e., principles that characterize the defining attributes of an ideal society. I use this model to show how we can interpret the components of Lipsey and Lancaster's mathematical model using concepts drawn from the familiar practice of analyzing hypothetical ideal societies to justify general normative principles. Because my model of normative reasoning represents a particular instance of the more general mathematical model, it

1 Lipsey and Lancaster note in passing that the theory of second best is "concerned with all maximization problems not just with welfare theory" (Lipsey and Lancaster, 1956, 12 note 2). It bears mentioning, however, that they promptly direct the reader's attention to additional economic applications to illustrate its breadth.

2 Compare Ingham (2019), which demonstrates a similar point about Arrow's "impossibility theorem".

exemplifies the second best theorem — it illustrates the implications of the general model for a particular domain.[3] We thus arrive at a novel *second best theorem for ideal normative theory*: roughly, if the ideal balance of two normative criteria (e.g., freedom and equality) is left unrealized, then we should not necessarily aim to realize the ideal balance for any remaining criteria (e.g., freedom and welfare, freedom and security, etc.).

While I extend the application of the theory of second best beyond applied normative reasoning, my model remains within the broader domain of normative social theory. Yet it would be a mistake to infer that the theory of second best applies only within this broader domain. The theory of second best is potentially relevant in any domain in which options are evaluated by aggregating disparate criteria, and trade-offs among these criteria are inescapable — philosophy of science (theory choice and scientific progress), formal epistemology (epistemic rules for nonideal agents), and computer science (optimizing algorithm performance) come to mind. My model of normative reasoning presents a template for extending the theorem to domains beyond normative social theory, although I leave it to others to develop these applications. The template is straightforward: construct a model of a typical mode of reasoning in some domain D that exemplifies the mathematical model underlying the theory of second best, and use this domain-specific model to develop an interpretation of the more general model using concepts from D. The result is a second best theorem for D.

The main aim of this paper is to show that the theory of second best is more general than scholars think in that it can be shown to apply to a wider range of domains than extant applications suggest. But it is also more specific than political theorists tend to think. Political theorists frequently intimate that *any* constraint on realizing the ideal is sufficient to trigger the theorem's warning about surprising and unexpected deviations from the ideal. This reflects a significant misunderstanding. I conclude by showing how the second best's warning about unexpected deviations is limited to a set of specific conditions, which circumscribe the mode of reasoning to which the theorem applies, the nature of the constraint that triggers the theorem, and the nature of the challenge generated by the theorem. Of particular importance, I show that some types of constraints on satisfying ideal principles raise no worries about unexpected deviations from the ideal — indeed, under these constraints, the prescribed adjustments are exactly as we would expect.

3 I owe this notion of a "theorem-exemplification" to Räikkä (2000).

1. SECOND BEST IN POLITICAL THEORY: A CASE OF LIMITED HORIZONS

There are long-standing questions in political theory about the relationship between normative ideals and political practice. Ideals ostensibly propose a standard by which to judge political practice, yet they often uphold states of affairs that are so distant from the status quo that they seem irrelevant for our reasoning about what to do in the real world (cf. Goodin, 1995, 37–8).[4] The theory of second best registers an important caution to those who seek to bring ideals to bear on practical politics in a straightforward manner. The theory emerged as a counterpoint to the enthusiasm for free markets generated by the first fundamental theorem of welfare economics, which establishes that perfectly competitive markets produce Pareto efficient allocations of socioeconomic goods (e.g., Debreu, 1959). This result demonstrates that, under stringent conditions, competitive markets realize a normative ideal: an outcome in which everyone is better off than they would be otherwise and no one can be made better off without making someone else worse off. It was widely recognized, of course, that the conditions for a perfectly competitive market are rarely if ever instantiated in the real world. Early social welfare analysis suggested a tempting reply: the best strategy for realizing a social welfare maximum is to approximate a perfectly competitive market as nearly as possible. The theory of second best shows this approximation strategy to be naïve. Using the same mathematical techniques used to prove the first welfare theorem, Lipsey and Lancaster demonstrated quite generally that, when a distortion arises in one sector of the economy to prevent the achievement of Pareto efficiency in that sector, then we will not necessarily achieve a welfare maximum by pursuing Pareto efficiency in the remaining sectors. Put simply, approximating perfectly competitive markets is not necessarily an effective strategy for maximizing social welfare.

Lipsey and Lancaster's theorem ostensibly bears an intuitive lesson — "be wary of approximations" — which can be readily extrapolated to domains outside economic theory, and to normative political theory in particular.[5] Consider two simple cases to

4 These questions have been revived under the rubrics of "ideal theory vs. nonideal theory" (Stemplowska and Swift, 2012; Valentini, 2012, 2017), "political moralism vs. political realism" (Rossi and Sleat, 2014), and "feasibility in political theory" (Southwood, forthcoming).

5 Stated this way, the insight seems so intuitive that one might think the result can be established using simple intuition pumps rather than Lipsey and Lancaster's formal mathematical argument (e.g., Estlund, 2018, 261). In the next section, I will show that Lipsey and Lancaster in fact demonstrate something more nuanced: "be wary of approximation reasoning *in cases where it is intuitively tempting to reason in this way*". Since simple intuition pumps rely heavily on intuition, they are ill-suited to establish this more nuanced claim. Using intuition pumps, one can either present cases for which it is highly intuitive to think that approximation reasoning fails, but then it merely establishes the existence claim that there are some cases

start.

> [A] person who prefers red wine to white may prefer either to a mixture of the two, even though the mixture is, in an obvious descriptive sense, closer to the preferred red wine than pure white wine would be. (Sen, 2009, 16)

> Your ideal car, let us suppose, would be a new silver Rolls. But suppose the dealer tells you none is available. The point of the general theory of second best is this: it simply does not necessarily follow that a car that satisfied two out of your ideal car's three crucial characteristics is necessarily second best. You may prefer a one-year-old black Mercedes (a car unlike your ideal car in every respect) over a new silver Ford (which resembles your ideal car in two out of three respects). (Goodin, 1995, 53)

These are "theorem-exemplifications" insofar as they present determinate and familiar cases to illustrate an implication of the second best theorem: among options that fall short of the ideal, the best option is not necessarily the one that looks most like the ideal.[6]

These exemplifications also provide analogies that help extend this second best insight to normative political theory in an accessible manner. Goodin draws the parallel explicitly:

> The same [lesson from the car case] applies [...] to our social and political prescriptions. In the best of all possible worlds, we would like all of our ideals to be realized simultaneously. [...] Ideally, we would like to attain liberty, equality, fraternity and material prosperity, all at one and the same time; but the classic trio might prove sociologically feasible only under conditions of severe material scarcity. (Goodin, 1995, 53)

The theory of second best thus provides an answer to those who, like naïve free market enthusiasts, would bring our normative ideals to bear on practical politics in a straightforward manner: we should be wary about pursuing an ideal as far as possible when we cannot realize it completely.

where approximation reasoning is invalid (viz., those cases where approximation reasoning is not intuitively tempting); or it can present simple cases where it is intuitively tempting to think approximation reasoning is valid, but then it cannot rely on intuition alone to persuasively show that, contrary to our intuition, such reasoning fails in these cases. The advantage of Lipsey and Lancaster's formal model is that it can represent a class of cases for which approximation reasoning is intuitively tempting and then supplement and discipline our initial intuition using mathematical reasoning to establish something counterintuitive.

6 See Räikkä (2000, 214) for this notion of a "theorem-exemplification".

While the above cases extend the second best insight to normative political theory, they also suggest a way to restrict its application within that domain. As others have noted, these cases are distinctive in describing both the ideal scenario and departures from it in terms that are too superficial or too crude to parallel the concepts and criteria we use when articulating abstract normative principles. To wit, it is hardly sensible to think of color or model year as normatively significant criteria for judging deviations from the ideal case (see, e.g., Gilabert 2012, 46; Swift 2008, 376; cf. Räikkä 2000, 213–17). This observation leads Goodin to concede that "[p]roblems of second best arise particularly when descriptions are couched in terms of surface attributes rather than more directly in terms of the underlying sources of those values," where examples of "surface attributes" include institutional features such as "liberal democracy with a market economy, welfare safety net and open borders" (Goodin, 1995, 53 note 45). More generally, political theorists have concluded that the significance of the theory of second best depends on the manner in which we describe the normative problem and, further, that it is most relevant in those cases where we describe our options in terms of institutional or policy attributes rather than abstract normative ideals (Räikkä, 2000, esp. 214–18). Put simply, political theorists have restricted the significance of the theory of second best to applied questions about how to most effectively implement more abstract principles (see also Brennan and Pettit 2005, 260–61; Coram 1996; Tessman 2010, 812–13).[7] Importantly, it leaves untouched our reasoning about the content of normative ideals. Goodin puts the point well: "What we are indexing to socio-psycho-economic circumstance are not the fundamental values themselves but merely the best mechanisms for attaining as many of them as possible. Timeless truths, ideally ideal ideals, remain. All that has to go are context-free political prescriptions for realizing them" (Goodin, 1995, 56).

Political theorists' restriction of the second best lesson to applied questions can be further explained by the fact that, when enumerating its implications for normative political theory, they have focused exclusively on generalizing insights from economists' informal applications of the second best theorem, which emphasize questions about the most effective institutional and policy means to realizing a constant normative goal,

7 Some exceptions: Räikkä develops an application to deontic logic (2000, 209), but his example is not in fact an exemplification of the theorem (this should become clear following the analysis below in section 4). Heath (2013) draws inspiration from the theory of second best to develop a way to distinguish between "different levels of idealization at which normative principles can be formulated" (164). Wiens (2016) also allows that the theory of second best might have implications for more abstract normative reasoning, yet he does not enumerate these potential implications in detail. Indeed, he sets aside the question of whether our reasoning about principles adheres to the form of optimization reasoning presupposed by Lipsey and Lancaster's proof (2016, 134).

namely, maximizing social welfare.[8] Indeed, political theorists typically introduce the theory of second best as a "well-known result in welfare economics" (Räikkä 2000, 204; cf. Goodin 1995, 52). It thus seems natural to limit the significance of the theory of second best, as economists have done, to the domain of applied normative reasoning.

These limited horizons are a result of limited engagement with Lipsey and Lancaster's original presentation.[9] To grasp the point, notice that the original "general theory of second best" consists of an abstract mathematical model ("the model"), a theorem of that model ("the theorem"), and a mapping from the model's mathematical objects to concepts in economic theory ("an interpretation"). Political theorists typically leave aside the math and focus on the economic interpretation of the model. Their attempts to extrapolate insights from the theorem to political theory are thus restricted by the insights that can be drawn from a specific interpretation of that theorem rather than the general model itself. Any interpretation of a model is bound to be limited in what it can show and thus in the insights it can produce. So before we accede to the received wisdom, we should examine whether there are plausible interpretations of the mathematical model that can extend the reach of the second best theorem beyond the domain of applied normative theory.

2. THE ORIGINAL MODEL AND ITS INTERPRETATION

To pave the way for new interpretations of the second best theorem, we review Lipsey and Lancaster's original presentation of the theory of second best.[10] Our objective is to become comfortable enough with the abstract model to see Lipsey and Lancaster's interpretation as just one possible interpretation. By prying apart the model from its most well-known interpretation, we can begin to see alternative interpretations and thereby expand the reach of the theorem beyond the limited confines it has occupied thus far. In the next section, I develop a novel interpretation of the model that makes clear the

8 "There is no doubt but that a theory of second best is oriented toward problems of policy" (Davis and Whinston, 1967, 323). See also Boadway (2017).

9 Relatedly, Wiens (2016) identifies the ways in which political theorists have misinterpreted the phrase "Paretian conditions" in the informal statement of the theorem and shows how these misinterpretations have led to significant misunderstandings of the theorem's implications for political theory. My discussion in this article complements this point.

10 The following sections focus on reconstructing and reinterpreting the mathematical model Lipsey and Lancaster used to prove their second best theorem with the aim of uncovering new conceptual insights. Since I wish to avoid alienating political philosophers who are instinctively skeptical about the use of mathematical models in political philosophy, I briefly address this general skepticism in section 5. Readers who are tempted to dismiss the mathematics out of hand are directed to those remarks.

theorem's implications for more abstract forms of normative reasoning.

2.1. Lipsey and Lancaster's basic model. Lipsey and Lancaster's proof starts from a model of a generic constrained optimization problem, which consists of three highly abstract mathematical objects.

- A set of n variables, each of which can take any real number as a value (i.e., they are continuous). A vector $(x_1, \ldots, x_n)$ denotes a particular assignment of values, one for each variable. These variables are used to characterize the attributes of objects in a set X.
- An *objective function* $F(x_1, \ldots, x_n)$, which encodes a goal or aim to be maximized (or minimized). F takes a vector $(x_1, \ldots, x_n)$ as an input and delivers a real number as an output. The "best element" in the set X is characterized by the vector $(x_1^*, \ldots, x_n^*)$ that maximizes (minimizes) the objective function.
- A *constraint* $G(x_1, \ldots, x_n) = 0$, which encodes a limit on the joint realization of values for $x_1, \ldots, x_n$ and thus specifies the set X.[11]

The functions F and G are assumed to be continuous. To illustrate the significance of this point for our purposes, take two arbitrary options, $(x_1, \ldots, x_n)$ and $(x_1', \ldots, x_n)$, and let $x_1 - x_1'$ be close to zero so that the two options are identical except for a slight difference in the attribute measured by the first variable. By the definition of a continuous function, $F(x_1, \ldots, x_n) - F(x_1', \ldots, x_n)$ will be close to zero so that the "value" of the two options, as indicated by F, differs very slightly. In effect, then, Lipsey and Lancaster's model picks out a class of cases that have the following features: (a) we can transition from one option to another by making arbitrarily small changes to the options' attributes, and (b) small changes from one option to another results in small changes in the "value" of the options. These are the kinds of cases for which it is intuitively tempting to think that, short of a well-defined best element, we do best by approximating that best element. Lipsey and Lancaster had no reason to make this nuance explicit; most members of their intended audience were familiar enough with the mathematics to have have recognized it. Political theorists have missed this point, however, because they ignore the math.

To see why this neglected nuance matters, consider the following simple case.[12] Suppose you see a doctor for an ongoing health problem and the doctor prescribes two

11 Lipsey and Lancaster use Φ instead of G to denote the constraint function. I follow a norm of subsequent treatments in using G instead of Φ (e.g., Ng, 2004, chap. 9).

12 I owe this case to Dave Estlund.

different liquid medications: you are to simultaneously take one ounce of each every day. The doctor explains that the two medications interact to boost their respective medicinal properties. Worried that this interaction might produce negative side-effects if you do not follow the prescription strictly, you ask, "What happens if I take half an ounce of one medication with one ounce of the other? Will they interact in ways that cause trouble? Would I be better off taking neither medication?" "No worries," the doctor replies, "up to the one ounce point, more is always better than less for both of these medications. They work best when you take one ounce of each together, but you will be fine if you take less than one ounce of either." Feeling reassured, you leave the doctor's office confident that, short of taking the full dosage, you are fine to approximate the full dosage. What is surprising about Lipsey and Lancaster's theorem is that it implies that, under certain conditions, this kind of approximation reasoning can fail in cases that have an analogous structure to this medical case. Because our intuition tells us that approximation reasoning is valid in these kinds of cases, we cannot readily see how this form of reasoning can nonetheless fail in these cases. To see this subtle point, we require a form of reasoning that can supplement and discipline our intuitions about these cases. This is why the mathematical model is useful.

After stating their model, Lipsey and Lancaster do not linger to observe its full generality. Immediately upon its specification, they note that "[t]his is a formalisation of the typical choice situation in economic analysis" and go on to re-state the problem as that of identifying a "Paretian optimum" (26). Although there is nothing in this model that is specific to economic analysis, Lipsey and Lancaster present the model and associated theorem exclusively in terms of concepts from orthodox welfare economics. For instance: the informal discussion at the outset of the paper frames the problem in terms of general equilibrium theory and "the attainment of a Paretian optimum" (11); the scope of the theory of second best is settled by considering "the role of constraints in economic theory" (12); the examples used to illustrate the general theory revolve around problems of efficient allocation of economic goods and, in particular, optimal tariff and tax policies.

Lipsey and Lancaster thus focus readers' attention on a particular *interpretation* of a more general model.[13]

- The set of n variables is interpreted as indicating quantities of the n goods produced by an economy and thus available for consumption by individuals. (For certain purposes, these variables can be more precisely interpreted as represent-

13 My discussion of Lipsey and Lancaster's interpretation is indebted to Ng (2004, esp. chaps. 2 and 9). See also Hoff (2000).

ing the allocation of each good to each individual.) A vector $(x_1, \ldots, x_n)$ indicates a particular quantity for each of the n goods.

- The function $F(x_1, \ldots, x_n)$ is interpreted as a *social welfare function*, which associates each allocation of n goods with a real number that represents the total social welfare generated by a particular allocation.[14] The "social welfare maximum" is characterized by the vector $(x_1^*, \ldots, x_n^*)$ that maximizes F.
- The constraint $G(x_1, \ldots, x_n) = 0$ is interpreted as a *production constraint*, which specifies limits on the joint production of the n goods and, in particular, the opportunity costs associated with transforming one good into another.[15]

This interpretation focuses the reader's attention on a determinate (and, for economists, familiar) constrained optimization problem, namely, that of identifying an allocation of economic goods that maximizes social welfare subject to a constraint on the joint production of these goods. This is for good reason: an interpretation of the model renders determinate the theorem's relevance and its implications for a concrete class of problems.

For our purposes later in the paper, it will be worthwhile to clearly understand how Lipsey and Lancaster use their model to characterize a (first-best) "Paretian optimum" and, for comparison, a "second best". To this end, we must work through some of the mathematical analysis of the model. To avoid getting lost in extraneous details, we note that the key points of interest are Lipsey and Lancaster's interpretation of and comparison between the systems of "proportionality conditions" defined in (3) and (8). Everything else is merely a necessary means to explaining how we arrive at these conditions and their relation to the theorem.

2.2. Characterizing a "Paretian optimum". To identify the solution to an optimization problem of the sort considered here, the standard technique is to take the partial derivatives of the Lagrangean function[16] $\mathcal{L}$ with respect to each variable x_i:

$$\mathcal{L}(x_1, \ldots, x_n, \lambda) = F(x_1, \ldots, x_n) - \lambda G(x_1, \ldots, x_n). \tag{1}$$

14 Lipsey and Lancaster call their various exemplifications of F a "community welfare function" (19), a "community preference function" (22), and a "utility function" (27, 28).

15 Lipsey and Lancaster use the term "transformation function" (22, 28).

16 See any textbook on constrained optimization for discussion of the Langrangean function.

This procedure delivers a system of equations that indicates the first-order necessary conditions for a welfare-maximizing allocation of goods.

$$\begin{aligned} F_1 - \lambda G_1 &= 0 \\ &\vdots \\ F_i - \lambda G_i &= 0 \\ &\vdots \\ F_n - \lambda G_n &= 0, \end{aligned} \tag{2}$$

where $F_i = \frac{\partial F}{\partial x_i}(x_i)$ and $G_i = \frac{\partial G}{\partial x_i}(x_i)$ are, respectively, the partial derivatives of F and G with respect to x_i. Solving this system of equations identifies the vector of quantities $(x_1^*, \ldots, x_n^*)$ that maximizes social welfare.

To relate the welfare-maximizing solution to the Pareto criterion, we can eliminate the "Lagrange multiplier" λ from each condition in (2) to derive a system of "proportionality conditions":

$$\begin{aligned} \frac{F_1}{F_n} &= \frac{G_1}{G_n} \\ &\vdots \\ \frac{F_i}{F_n} &= \frac{G_i}{G_n} \\ &\vdots \\ \frac{F_{n-1}}{F_n} &= \frac{G_{n-1}}{G_n}, \end{aligned} \tag{3}$$

with x_n being chosen, without loss of generality, as a reference good for computing exchange rates (the "numeraire" in econ-speak). Given the interpretation of F as a social welfare function and G as a production function, we can interpret $\frac{F_i}{F_n}$ as the "marginal rate of substitution" (MRS) for goods i and n. The MRS is the rate at which consumers can exchange a small amount of good i for a given amount of good n while maintaining a constant level of social welfare. Analogously, we can interpret $\frac{G_i}{G_n}$ as the "marginal rate of transformation" (MRT) for goods i and n. The MRT is the rate at which producers can redirect production of a small amount of i toward production of a given amount of good n; alternatively, it represents the opportunity cost of producing more of good n in terms of the amount of good i that would be given up. The proportionality conditions in (3) thus indicate that a welfare-maximizing allocation of goods must equalize the MRS and MRT for every pair of goods, which is also a necessary condition for a Pareto efficient

allocation (see, e.g., Ng, 2004, chap. 2).

2.3. Characterizing a "second best" optimum. The conditions in (3) state what Lipsey and Lancaster call the "Paretian optimum conditions"; these are the conditions used to characterize a first-best Pareto optimal allocation of goods. In their model, a second best problem arises when one of these proportionality conditions is unsatisfied for one reason or another. To represent such a problem, we assume (without loss of generality) that the MRS for goods 1 and n is not equal to the MRT for the same pair:

$$\frac{F_1}{F_n} = k\frac{G_1}{G_n} \text{ (for arbitrary } k \neq 1). \tag{4}$$

To characterize the "second best" allocation — that is, the welfare-maximizing vector subject to this new constraint — we use the same procedure as above: we first take the partial derivatives of the amended Lagrangian function $\mathcal{L}'$, which adds the new second-best constraint to our original function:

$$\mathcal{L}'(x_1,\ldots,x_n,\beta,\mu) = F(x_1,\ldots,x_n) - \beta G(x_1,\ldots,x_n) - \mu\left(\frac{F_1}{F_n} - k\frac{G_1}{G_n}\right). \tag{5}$$

This gives us a new system of first-order conditions, which we can solve to identify the second-best social welfare maximum:

$$\begin{aligned} F_1 - \beta G_1 - \mu H_1 &= 0 \\ &\vdots \\ F_i - \lambda G_i - \mu H_i &= 0 \\ &\vdots \\ F_n - \lambda G_n - \mu H_n &= 0, \end{aligned} \tag{6}$$

where H_i stands in for a more complicated expression:

$$H_i \equiv \frac{F_n F_{1i} - F_1 F_{ni}}{F_n^2} - k\frac{G_n G_{1i} - G_1 G_{ni}}{G_n^2}. \tag{7}$$

Although its complexity is central to Lipsey and Lancaster's proof, we can leave aside a detailed exposition of H_i here. Two points are relevant. The first is that the conditions in (6) imply a new system of proportionality conditions for a second-best welfare-maximizing allocation of goods[17]:

17 Note that Lipsey and Lancaster's original presentation of these conditions (their 7.6) contains a typo-

$$
\begin{aligned}
\frac{F_1}{F_n} &= \frac{G_1}{G_n}\left[\frac{1+\frac{\mu}{\beta G_1}H_1}{1+\frac{\mu}{\beta G_n}H_n}\right] \\
&\vdots \\
\frac{F_i}{F_n} &= \frac{G_i}{G_n}\left[\frac{1+\frac{\mu}{\beta G_i}H_i}{1+\frac{\mu}{\beta G_n}H_n}\right] \\
&\vdots \\
\frac{F_{n-1}}{F_n} &= \frac{G_{n-1}}{G_n}\left[\frac{1+\frac{\mu}{\beta G_{n-1}}H_{n-1}}{1+\frac{\mu}{\beta G_n}H_n}\right]
\end{aligned}
\tag{8}
$$

The second point is that the first-best proportionality condition in (3) for a pair $\langle i, n\rangle$ does not match the second-best proportionality condition in (8) for the same pair whenever the expression in the square brackets is not equal to 1. To see the significance of this point, note that we already know that the second best will not satisfy the "Paretian optimum condition" in (3) for the pair $\langle 1, n\rangle$; we assumed this in (4) to generate the second best problem. The question is whether the second best allocation nonetheless approximates the "Paretian optimum" by satisfying the proportionality conditions in (3) for the remaining unconstrained pairs. The comparison between (3) and (8) points to our answer: a second best allocation does not satisfy the proportionality condition for a Pareto optimum for a pair $\langle i, n\rangle$ if the expression in the square brackets is not equal to 1 for that pair.

The crux of the theory of second best is that the expressions in the square brackets in (8) are not necessarily equal to 1 and, thus, the second-best proportionality conditions are not necessarily equivalent to the first-best "Paretian conditions" (Lipsey and Lancaster, 1956, 27–8). Subsequent analyses have shown that the conditions in (8) match the conditions in (3) if and only if the functions F and G satisfy certain *separability conditions* (Davis and Whinston, 1965; Jewitt, 1981; Blackorby, Davidson and Schworm, 1991). An explanation of these separability conditions is outside the scope of this article (but see Ng, 2004, 195–96). What's important here is that separability is a very restrictive assumption — it would be surprising if it held in an economy with "thousands of products and inputs" (Ng, 2004, 195) and "many [heterogeneous] consumers" (Blackorby, Davidson and Schworm, 1991, 269). In the context of economic theory, the case in which the second-best conditions in (8) match the first-best conditions in (3) is thus a "special case" (Ng,

graphical error, which is corrected in Lipsey and Lancaster (1997).

2004, 195).

2.4. A second best theorem for economic theory. Thus, we arrive at Lipsey and Lancaster's theorem. To quote them:

> [I]f there is introduced into a general equilibrium system a constraint which prevents the attainment of one of the Paretian conditions [in (3)], the other Paretian conditions, although still attainable, are, in general, no longer desirable. (Lipsey and Lancaster, 1956, 12)

We can restate this in a way that enables us to expose the two key lessons of this section.

> A welfare-maximizing Pareto optimal allocation of goods requires that the MRS and MRT be equalized for every pair of goods in the economy, as characterized by (3). Suppose that the MRS and MRT cannot be equalized for one pair of goods, as in expression (4). Then it is usually the case (i.e., when stringent separability assumptions fail to obtain) that a welfare-maximizing second best allocation, which is characterized by (8), will fail to equalize the MRS and MRT for the remaining pairs of goods. (cf. Lipsey and Lancaster, 1956, 26)

The first lesson is that all the action in the theorem is in the representation of a second best problem in expression (4) and the characterization of first-best and second-best solutions using the conditions in (3) and (8) respectively.

The second lesson is that the familiar applications of the theorem to social welfare analysis follow from Lipsey and Lancaster's specific interpretation of these mathematical expressions — in particular, their interpretation of the terms in (3) as consumers' marginal rates of substitution and producers' marginal rates of transformation. None of these mathematical expressions is essentially about the substitution and transformation of economic goods. The theorem is, as Lipsey and Lancaster note in passing, "concerned with all maximization problems not just with welfare theory" (Lipsey and Lancaster, 1956, 12 note 2). It is, in the first instance, a theorem about mathematical optimization in general. It is easy to lose sight of this fact because the theorem has been applied almost exclusively to the domains of economic theory and welfare policy. But we must be clear that the theorem's implications for these domains are wholly due to the plausibility of a specific interpretation of the more general mathematical model.

3. EXPANDING THE REACH OF THE SECOND BEST THEOREM

Drawing together the insights of the previous section, we can see how to extend the range of the theorem beyond its current reach — namely, by showing how the mathematical model admits of plausible interpretations beyond those adopted by economists. Here I focus on expanding the theorem's reach in normative political theory.[18] To this end, I develop an interpretation of the model that shows how the theorem can have important implications not only for applied questions about how to implement normative ideals but also for more abstract questions about the content of the ideals we should try to implement. I develop this interpretation in three steps (which parallel the structure of my reconstruction in section 2). First, I show how a relatively widespread form of abstract normative reasoning can be modeled by a more general constrained optimization problem. Second, I show that this form of normative reasoning admits the existence of a "first best scenario" that can be plausibly represented using the expressions in (3). Third, I show that this form of normative reasoning admits of a "second best problem" that can be plausibly represented using the expression in (4) and, thus, of a "second best scenario" that can be plausibly represented by the expressions in (8).

The economic interpretation of the mathematical model is readily accepted because it is familar — few social scientists balk at the idea of using a mathematical function to represent the social welfare of our options. In contrast, the normative-theoretic interpretation I wish to propose is unfamiliar and thus liable to be rejected solely on that account. To enhance the plausibility of my proposed interpretation, I adopt a strategy of *exemplification*: I develop a thought experiment that is simultaneously (i) an instance of the Lipsey and Lancaster model and (ii) an abstract representation of a form of normative reasoning that political theorists use to specify the content of general principles that are meant to characterize normatively ideal societies. What I have in mind for (ii) is a form of comparative reasoning about hypothetical societies that is plausibly used by (e.g.) John Rawls (1999) to justify the claim that his principles of justice characterize an ideally just society, or by G.A. Cohen (2009) to justify the claim that his two socialist principles characterize a normatively ideal society. I do not pursue these claims here, nor any claims about how widely this form of reasoning is used in practice; this is beyond the scope of this paper. It will be enough to achieve my aims if I succeed in presenting a model of normative reasoning that is recognizable as a form of reasoning that some theorists

18 Several people have suggested to me that it would be worthwhile to develop applications to other domains such as theory choice in science or formal epistemology. I leave it to others to develop these applications; I hope my discussion here will provide a useful template for this work.

plausibly use in practice to specify the content of general normative principles rather than the institutions or policies that best implement more abstract principles.[19]

To forestall certain distractions, we should be clear that my exposition of this thought experiment will present the target form of reasoning about principles as being much more precise than it is in practice. In practice, our reasoning about the principles that characterize an ideal society is an imprecise and sometimes messy affair. But my aim here is not to faithfully render all facets of this type of reasoning as it appears in practice, warts and all. My aim is instead to present a simplification of this type of reasoning that faithfully represents its core logical structure so that we can better see important implications of the reasoning we often use to specify normative principles. For my purposes, the model is supposed to capture a mode of reasoning whereby comparisons among possible societies are made by reference to several abstract normative criteria (e.g., freedom, equality, welfare, and so on), and ideal normative principles are specified by analyzing certain defining attributes of the possible society that fares best with respect to these comparative judgments. The precision imposed by the mathematics is simply a way to discipline our efforts to draw conclusions from this thought experiment.

3.1. The basic model. Imagine we are evaluating possible social arrangements ("societies"). We describe possible societies in terms of three normatively significant criteria: *freedom*, *equality*, and *security*. To make things concrete, let's suppose that we understand freedom in negative terms and operationalize it using a variable f that measures the extent to which people are de facto permitted to act as they see fit unrestrained by government coercion (for simplicity, we assume that each member of society experiences the same degree of freedom).[20] We understand security in terms of law and order and operationalize it using a variable s that measures the extent to which (e.g.) individuals' physical integrity is assured, the performance of contracts is assured, and so on. Finally, we understand social equality in material terms and operationalize it using a variable e that measures the statistical distribution of income and wealth, such as the Gini co-

19 Compare Aristotle: "Our purpose is to consider what form of political community is best of all for those who are most able to realize their ideal of life. We must therefore examine not only this but other constitutions, both such as actually exist in well-governed states, and any theoretical forms which are held in esteem, so that what is good and useful may be brought to light" (1988, 20–1). Or, more recently, Simmons: we specify principles of ideal justice by "compar[ing] the operation of societies ordered by competing principles of justice while assuming strict compliance with those principles" (2010, 8).

20 We do not claim that this is the best or even a desirable way to conceptualize freedom, although it is clearly recognizable as a conceptualization that some people find attractive. Our chief aim is to construct an intuitively tractable thought experiment. The same caveat applies for our conceptualizations of equality and security.

efficient. For simplicity, we treat each variable as a continuous measure ranging from 0 (complete absence) to 1 (full realization). (Since a Gini coefficient of $g = 1$ implies complete inequality, we measure equality as $e = 1 - g$ to ensure that $e = 1$ indicates full equality.)

We wish to characterize the ideal society, which we identify with the (hypothetical) society that has the highest overall normative value. Assume that we judge overall normative value solely as a function of the extent to which a society realizes our three criteria. For simplicity, we assume that, according to this function, the overall value of a society increases as the realization of each criterion increases. We also assume that each of the three criteria are necessary conditions for a society to have any value and that the three criteria are given equal weight in our evaluative judgments. More formally, we assume that we comparatively evaluate possible societies according to a value function[21]

$$F(e, f, s) = e \times f \times s. \tag{9}$$

Suppose, finally, that there are limitations on the joint realization of freedom, equality, and security. Since we are concerned to identify the ideal society, we might think of these limitations as akin to the kinds of assumptions ideal theorists are willing to make about the background circumstances in which we are to establish our society (e.g., Rawlsian circumstances of justice). These limitations can be as idealized as we like. Indeed, we can abstract from empirical circumstances entirely — we can conceptualize these limitations as being due to inherent conflicts among the concepts. The only restriction is that we cannot fully realize all three variables simultaneously. We conceptualize these limitations as opportunity costs; for example, that increasing the realization of freedom entails decreasing the realization of security or equality (or both) by some amount. More concretely, we assume that these opportunity costs satisfy the following condition: starting from full freedom, small decreases in freedom bring large gains in security (or equality), but as we approach full security (equality), small increases in security (equality) require increasingly

21 Although we choose a specific functional form for F, we make no normative claims here; in particular, we do not claim that this is the only, or even the best, function for determining the value of possible societies. Recall that our objective is to develop a determinate model of normative reasoning that exemplifies the implications of Lipsey and Lancaster's more general model. We thus need a function F that: (i) takes at least three variables as arguments; (ii) is a continuous function (see my remarks in section 2.1). Additionally, we specify a functional form that is not additively separable (since, as noted above, separability is sufficient to skirt the theorem). Our chosen functional form is among the simplest instances of a value function that is sufficient to exemplify the implications of the second best theorem and is independently plausible on normative grounds. But one is free to use a different value function here, so long as it satisfies these conditions.

large decreases in freedom. We formalize these conditions using a "realization function"

$$G(e, f, s) = -1 + f + e^2 \times s^2 \tag{10}$$

and, following convention, we set this constraint equal to 0 (denoted $G = 0$ for short).

We now have the materials to interpret the components of the general optimization model in terms of abstract normative reasoning.

- We interpret our options as possible (hypothetical) societies, and we interpret the set of variables used to describe these societies as three normatively significant criteria.
- We interpret the objective function F as an *overall normative value function*, which encodes the manner in which our three criteria jointly affect our normative evaluations using the expression in (9).
- We interpret the constraint function G as a *realization function*, which encodes limitations on the joint realization of our three criteria using the expression in (10).

3.2. Characterizing the ideal society. Given our model of normative reasoning, we can think of the ideal society as the (hypothetical) society that maximizes overall normative value as indicated by F subject to the constraint indicated by $G = 0$.[22] Following our discussion in section 2, we proceed by taking the partial derivatives of the function

$$\mathcal{L}(e, f, s, \lambda) = efs - \lambda(-1 + f + e^2 s^2), \tag{11}$$

which is an instance of the expression in (1). This implies the following first-order necessary conditions for a value-maximizing society:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial e} &= fs - \lambda 2es^2 = 0 \\ \frac{\partial \mathcal{L}}{\partial f} &= es - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial s} &= fe - \lambda 2se^2 = 0 \end{aligned} \tag{12}$$

22 Although the details to follow depend on particular functional forms, the conceptual point we wish to convey does not, since this is an exemplification of a more general model. In other words, we could have specified any particular functional forms, so long as they are instances of the type of functions the constitute the general mathematical model. This is made especially clear in an appendix.

Solving this system of equations, we find that our ideal society realizes our freedom, equality, and security variables at the following levels: $e = s \approx 0.76$ and $f \approx 0.67$. This gives us one way to characterize the ideal society's defining attributes, namely, in terms of the extent to which it realizes our three criteria.

Additionally, we can characterize our ideal society by rearranging the first-order conditions to derive two equalities:

$$\frac{f}{s} = 2se^2 \qquad\qquad \frac{f}{e} = 2es^2 \tag{13}$$

We can treat these equalities as indicating the relative significance granted to our three normative criteria in terms of ratios. For example, at the ideal society, the ratio of freedom to equality is equal to two times the product of the level of equality and the squared level of security.[23] So a second way to characterize our ideal society's defining attributes is in terms of the relative significance assigned to our three criteria at the ideal society.

To bring our model closer to conventional normative theorizing, notice that these two ways of characterizing the defining attributes of an ideal society plausibly perform a similar function to that performed by two familiar kinds of normative principle. The first kind, which I shall call *level principles*, specifies the level of each variable to be realized at the ideal society. An example of such a principle in our model is "Freedom (conceptualized as above) should be realized at the rate of $\frac{2}{3}$ of full freedom", or $f \approx 0.67$. This is functionally analogous to more familiar normative principles such as "Every citizen's basic needs should be satisfied" or "All citizens should have equal influence on political decisions". The second kind of principle, which I shall call *ratio principles*, specifies the ideal balance of our three criteria. An example of such a principle in our model is "A freedom-to-equality ratio that is equal to two times the product of the level of equality and the squared level of security should be realized", or $\frac{f}{e} = 2es^2$. This is functionally analogous to principles such as "Citizens' basic material needs should be satisfied before we turn our attention to equalizing political rights" or "Granting every citizen a share of the means of production is more important than maximizing total social welfare".[24]

No doubt, the principles in our model are much more precise than familiar normative

23 Notice that, since $e = s \approx 0.7598$ and $f = \frac{2}{3}$, $\frac{f}{e} \approx 0.877$ and $2es^2 \approx 0.877$.

24 If it seems less plausible that conventional normative theorizing involves ratio principles, consider Rawls's remarks on intuitionism and the problem of constructing indifference curves by assigning weights to a plurality of values. Much of Rawls's theory is motivated by a desire to say something systematic about the relative significance of disparate values (see, e.g., Rawls, 1999, secs. 8, 9). Swift (2008, 369) makes a similar point.

principles, but this disparity in precision is beside the point here. Two points matter for our purposes. The first is that our model of normative reasoning contains elements that we can plausibly interpret as performing the same function as more familiar normative principles. The second is that, like the model principles, we can plausibly think of the more familiar principles as deriving from an estimation (although a rough and ready one) of how to best balance several different normative considerations given the opportunity costs of their joint realization. Thus, we have the resources to interpret the mathematical characterization of a "first-best scenario" in terms of ideal normative principles.

- We interpret the solution to the first-order necessary conditions in (12) as *ideal level principles*, which indicate the extent to which we (ideally) ought to realize each normative criterion.
- We interpret the proportionality conditions in (13) as *ideal ratio principles*, which indicate the ideal balance among our normative criteria.

In an appendix, I show that the ideal society as characterized in our model satisfies a third, more abstract *meta-principle*, which constrains the content of ratio principles. Here's a rough statement of this principle:

> The normatively ideal balance between two criteria x and y (i.e., the content of the ratio principle that governs x and y) is determined by the relative opportunity costs of their respective realization.

This meta-principle is the analogue to Lipsey and Lancaster's "Paretian conditions", which require that the consumers' marginal rate of substitution for two goods i and j must be equal to producers' marginal rate of transformation for i and j (see the discussion at the end of section 2.2). A full understanding of the reasoning behind this meta-principle is useful for making clear how our model of normative reasoning is an instance of Lipsey and Lancaster's more general model. Unfortunately, it requires a deep dive into certain mathematical details (viz., the substantive meaning of partial derivatives), and my effort to interpret these details in an accessible way is liable to distract from the main thread of this section; hence the appendix. Readers who wish to see our model in this section explicitly linked to the more general mathematical model are encouraged to work through the appendix. We will return to this meta-principle in section 4.1.

3.3. Characterizing the second best society. Now that we have characterized an ideal society in terms of general normative principles, the question prompted by the theory of second best is, how (if at all) do principles for a second best society deviate from the ideal

principles? More precisely: If we are constrained to leave an ideal principle unsatisfied, would a second best society nonetheless approximate the ideal by satisfying the remaining ideal principles? To answer this question, we need to characterize the constraint that triggers the question, as well as a way to characterize the defining attributes of a second best society — that is, the society that has the highest overall normative value given the new constraint.

Recall our discussion in section 2.3: the second best theorem is triggered when a constraint of the sort specified in expression (4) is imposed on our normative reasoning. The ratio principles that characterize the ideal — the expressions in (13) — are instances of the proportionality conditions in (3) (see appendix). Thus, characterizing a second best society becomes a concern when we are constrained to leave a single ratio principle unsatisfied.[25]

For concreteness, suppose that we are constrained to leave the ideal freedom-to-security ratio principle unsatisfied, such that

$$\frac{f}{s} = 2(2se^2), \tag{14}$$

where k from expression (4) is set equal to 2. So we are constrained to realize freedom at four times the rate of the product of security and equality squared; in comparison with the ideal society, we are constrained to under provide security and equality relative to freedom. Importantly for our purposes, nothing prevents us from satisfying the ideal freedom-to-equality ratio principle, nor any of the three level principles (taken separately).

To characterize the second best society given this new constraint, we take the partial derivatives of the function

$$\mathcal{L}'(e,f,s,\beta,\mu) = efs - \beta(-1+f+e^2s^2) - \mu\left(\frac{f}{s} - 4se^2\right), \tag{15}$$

which is the same as expression (11) with the addition of the new constraint term (it is also an instance of (5)). The new first-order necessary conditions for a second best society

25 I use this somewhat cumbersome phrase to avoid suggesting that a second best constraint must always *prevent* us from realizing the ideal. Räikkä (2000, 211f) rightly points out that theory of second best is not necessarily about what to do when the ideal is infeasible, which Lipsey (2007) also recognizes.

are

$$\begin{aligned}
\frac{\partial \mathcal{L}'}{\partial e} &= fs - \beta 2es^2 - \mu(-8es) = 0 \\
\frac{\partial \mathcal{L}'}{\partial f} &= es - \beta - \mu\left(\frac{1}{s}\right) = 0 \\
\frac{\partial \mathcal{L}'}{\partial s} &= fe - \beta 2se^2 - \mu\left(\frac{-f}{s^2} - 4e^2\right) = 0
\end{aligned} \tag{16}$$

Solving this, we find that there is no unique second best society but instead numerous societies that are equal second best. Every second best society realizes $f = 0.8$. Additionally, a society that realizes $f = 0.8$ is second best if (and only if) it realizes a combination of equality and security such that $e \times s \approx 0.45$; some examples include $(e = 0.5, s \approx .89)$ and $(e = 0.75, s \approx 0.6)$.[26] We can see immediately that, in light of the constraint on realizing the ideal freedom-to-security ratio principle, a second best society departs from the ideal freedom level principle. Additionally, while there is a second best society that continues to realize the ideal equality level principle, this second best society must depart from the ideal security level principle. (A similar point holds for a second best society that realizes the ideal security level principle.) Beyond these cases, however, there are many equal second best societies that depart from the ideal with respect to all three level principles. In any case, this departure from the ideal is not where we want to focus our attention.

By assumption, we are constrained to deviate from the ideal freedom-to-security ratio principle; this triggers the second best problem. Should we nonetheless satisfy the ideal freedom-to-equality ratio principle, assuming that this is not constrained? To answer this question, start by rearranging our first-order conditions in (16) to derive the ratio principles that characterize the second best society:

$$\frac{f}{s} = \frac{2se^2}{1}\left[\frac{1 + \frac{\mu}{\beta 2se^2}\left(\frac{-f}{s^2} - 4e^2\right)}{1 + \frac{\mu}{\beta}\left(\frac{1}{s}\right)}\right] \qquad \frac{f}{e} = \frac{2es^2}{1}\left[\frac{1 + \frac{\mu}{\beta 2es^2}(-8es)}{1 + \frac{\mu}{\beta}\left(\frac{1}{s}\right)}\right] \tag{17}$$

A detailed interpretation of these conditions is not relevant here; the important point is that these are instances of the more general conditions in expression (8) (see appendix). Following our present interpretation of the general model, the left-hand equality represents the freedom-to-security ratio principle that characterizes the second best society,

26 Precisely, a society maximizes normative value given the second best constraint if and only if $f = \frac{4}{5}$ and $es = \frac{1}{\sqrt{5}}$. A second best society realizes at least $e \approx 0.45$ (when $s = 1$) or $s \approx 0.45$ (when $e = 1$).

while the right-hand equality represents the freedom-to-equality ratio principle that characterizes the second best society.

By assumption, the second best freedom-to-security ratio principle deviates from the ideal, so we know that the bracketed term for the left-hand equality is not equal to 1 (in fact, by expression (14), it must be equal to 2). The question here is whether the freedom-to-equality ratio that characterizes the second best society deviates from the ideal ratio. Following our discussion in section 2.3, the second best society is characterized by the ideal freedom-to-equality ratio principle only if

$$\frac{1+\frac{\mu}{\beta 2es^2}(-8es)}{1+\frac{\mu}{\beta}\left(\frac{1}{s}\right)}=1. \tag{18}$$

This equality holds only if $\mu = 0$, but this can't be true.[27] If $\mu = 0$, then it follows that the second best constraint in expression (14) does not bind. In other words, if we are constrained to leave the ideal freedom-to-security ratio principle unsatisfied, then $\mu \neq 0$ by assumption.[28] So we must assume that $\mu \neq 0$, which implies that the second best society does not satisfy the ideal freedom-to-equality ratio principle. Thus, if we are constrained to leave the ideal freedom-to-security ratio principle unsatisfied, then, if we want to realize a second best society, we ought not approximate the ideal by satisfying the ideal freedom-to-equality ratio principle.

3.4. A second best theorem for ideal theory. In section 3.1, we established an interpretation of Lipsey and Lancaster's mathematical model in terms an abstract form of normative reasoning. We imagined that we evaluate possible social arrangements as a function of the extent to which they realize three normatively significant criteria. We then identified the ideal society with the (hypothetical) society that ranks highest given a constraint on the joint realization of our three criteria. Importantly, we did not assume that this constraint encodes realistic social conditions, but allowed it to encode highly idealized social conditions that are consistent with the assumption that all three criteria cannot be fully realized simultaneously. Given this set up, we turned in section 3.2 to specifying certain defining attributes of the ideal society. This exercise enabled us to characterize the ideal society using two kinds of normative principles — namely, level principles, which specify the extent to which each of the criteria is realized, and ratio

27 The expression is undefined if $e = 0$ or $s = 0$.

28 If this explanation is unsatisfying, notice that if $\mu = 0$, then the bracketed term for the left-hand equality in (17) will be equal to 1, which implies — contrary to our assumption — that the second best society satisfies the ideal freedom-to-security ratio principle.

principles, which specify the ideal balance of these criteria. With our characterization of the ideal in hand, we turned in section 3.3 to characterizing a second best society. We started by noting that the constraint that triggers the exercise applies specifically to the realization of an ideal ratio principle. We then specifed the defining attributes of a second best society given this new constraint. The central result is a derivation of the ratio principles that characterize a second best society, specified in expression (17).

The comparison between the ideal ratio principles in (13) and the second best ratio principles in (17) leads us to the following conclusion for this particular model:

> If we are constrained to leave the ideal freedom-to-security ratio principle unrealized, then we should not aim to satisfy the ideal freedom-to-equality ratio principle even if nothing prevents us from doing so.

All of the mathematical expressions in our model are instances of the expressions that constitute the more general model in section 2 (see appendix). Thus, a generalization of this conclusion holds as a *second best theorem for ideal normative theory*.

> Suppose we evaluate possible societies as a function of the extent to which they realize certain normatively salient criteria, and suppose that the ideal is characterized by (among other things) principles that specify the ideal balance of these criteria. If we are constrained to leave one ideal ratio principle unsatisfied, then we should not necessarily satisfy the remaining ideal ratio principles even if we can do so.

4. CONSTRAINING THE THEOREM'S SIGNIFICANCE

We have shown that the general theory of second best is more general than is often supposed by extending its reach to a new domain. By developing a plausible exemplification of the general model, we have also shown how additional interpretations might be developed and the theorem's reach further extended beyond its conventional home in applied social theory. Yet the theorem's significance within any domain depends on the specific conditions that limit its application, conditions such as the specific type of constraint that triggers the warning against approximating the ideal. In this section, we deepen our understanding of the theorem's significance for ideal normative theory by examining the conditions that circumscribe its implications within this domain.

We start by noticing that our model of normative reasoning in section 3 presupposes that we specify the defining attributes of an ideal society by comparatively evaluating possible social arrangements in terms of more basic normative criteria. This, in turn,

presupposes that we have analyzed the normative concepts implicated by these criteria (e.g., freedom, equality, security), and have specified a value function that aggregates these criteria for the purposes of normatively evaluating possible social arrangements. Our model of reasoning about the defining attributes of the ideal society thus presupposes something akin to G.A. Cohen's "fundamental normative principles", which define "a set of [...] indifference curves whose axes display packages of different extents to which competing principles [i.e., criteria] are implemented" (Cohen, 2003, 245).[29] Our reasoning about these items need not be as precise as the mathematical model seems to suggest; it need only be the case that our reasoning about the principles that characterize an ideal society conform to the same logical structure — namely, that ideal principles are a function of some more basic criteria (formalized here as F), which we use to comparatively evaluate possible societies. Thus, to apply the theory of second best to our reasoning about ideal societies, we must have already engaged in some normative reasoning at a higher order of abstraction.

Comparing the presuppositions of our model to Cohen's fundamental principles might suggest two further limits on our application of the second best theorem to ideal theory: that the theorem is irrelevant for our reasoning about fundamental principles, and that the theorem only applies to our reasoning about "principles of regulation". The first point is too hasty, the second is mistaken. To the first point: Notice that any interpretation of the general mathematical model, if it is to be sufficiently determinate, must limit the application of the theorem to a particular class of cases. We do not extend the theorem's reach by developing a "master interpretation" that covers all relevant cases. Instead, we develop a plurality of interpretations, each of which identifies a determinate subset of the cases to which the theorem applies. In this paper, we have developed an interpretation that shows how the theorem applies to our reasoning about the defining attributes of an ideal society. Whether our reasoning about fundamental normative principles is also subject to a second best theorem depends on whether we can develop a plausible model of our reasoning about such principles that conforms to the logical structure of a general optimization problem.

To the second point: Recall that, for Cohen, a principle of regulation specifies "a certain type of social instrument, to be legislated and implemented" (Cohen, 2003, 241); it is, in other words, a principle for specifying the institutions, practices, and policies that should be implemented as a means to realizing some balance of more basic values and principles. But our model of ideal theoretic reasoning does not characterize ideal societies

29 Note that Cohen asserts that these indifference curves are "fact-independent". I think we can remain noncommittal on this point here.

in terms of institutions, practices, and policies, but rather in terms of normative principles that specify in a general and abstract way the appropriate balance of these more basic principles. Unlike Cohen's principles of regulation, the level and ratio principles we use to characterize the ideal society do not prescribe any particular institutional scheme or policy program; they do not specify any particular "device for having certain effects" (Cohen, 2003, 241). Instead, they specify certain effects — i.e., a normatively ideal balance of more basic criteria — that are to be realized by institutional design and policy choice. Thus, we have modeled a form of normative reasoning that operates at a middle-level of abstraction, below our reasoning about basic normative criteria but above our reasoning about institutional design and policy choice.[30] The upshot of the preceding points is that the second best theorem in the previous section is limited to a particular form of normative reasoning, namely, reasoning that specifies the content of general middle-level normative principles by comparatively evaluating possible societies in terms of the extent to which they realize certain basic normative criteria.

Within the specified domain, the significance of the second best theorem is further limited to the imposition of a specific type of constraint. This is a subtle point, and one that is easy to miss if one relies on informal glosses of the theorem. It is most common for political theorists to claim that the second best theorem is relevant when we shift our attention from idealistic circumstances to more realistic social circumstances.[31] This shift from ideal to nonideal circumstances can in principle generate several different types of constraints on the realization of ideal principles. Judging from the typical gloss, any of these constraints is sufficient to trigger a concern that nonideal prescriptions will deviate from ideal principles in surprising and unexpected ways — the second best theorem is often said to apply when vague "ideal conditions" or "desiderata" are left unsatisfied.[32] But this would be a serious misunderstanding of the theorem's significance for normative theory. The theorem's warning is more specific: that we should not necessarily approximate the ideal *by satisfying as many ideal ratio principles as possible* — that nonideal ratio principles will likely deviate from ideal ratio principles in unexpected ways. Moreover, this warning is not triggered by just any sort of constraint on the realization of ideal principles, but specifically by a constraint on satisfying an ideal ratio principle. To see the

30 Compare Sangiovanni's discussion of "middle-level, or *mediating*, principles", which occupy a level of abstraction between Cohen's fundamental principles and principles of regulation (2016, 15, original emphasis; cf. 14–6, 19–20).

31 See Wiens's discussion of the "background assumptions" interpretation of the theory of second best and the citations therein (2016, 137–41).

32 Consider, for example, Estlund's gloss: "When there are several desiderata that are desirable as a package, if one of them is not satisfied, the value of the rest of them is thrown back into question" (2011, 216).

significance of this point and thus enhance our understanding of the theorem's significance for normative theory, we can use our model to consider various additional types of constraints and show that these do not generate any serious concerns about unexpected deviations from ideal principles.

We have already considered one type of constraint on realizing ideal principles: in section 3.3, we stipulated a constraint on satisfying an ideal ratio principle and showed that this leads to a second best theorem for ideal theory. There are two additional types of constraint that we can impose within our model: a shift to nonideal circumstances could lead to a change in the opportunity costs to realizing our normative criteria, or it could lead to a constraint on the realization of an ideal level principle. We treat each of these in turn.[33]

4.1. Changing opportunity costs. Recall that the realization constraint $G = 0$ in expression (10) is meant to represent nothing more than an assumption that we cannot fully realize all of our normative criteria at the same time. We make no assumptions about the reason for these costs and, importantly, we allow that they may be due to inherent conflicts among the concepts rather than any unfavorable empirical conditions. Imagine that we have completed our ideal theory assuming these ideal opportunity costs. Our ideal theory prescribes the level principles $f \approx 0.67$ and $e = s \approx 0.76$ and the ratio principles in (13) (recall section 3.2). Suppose now that a shift to nonideal circumstances requires us to adjust our estimate of the relevant opportunity costs. For concreteness, suppose that

$$G(e, f, s) = -1 + f + 2e^2 s^2 = 0, \tag{19}$$

which implies that increasing the realization of security or equality imposes a greater sacrifice of freedom (i.e., has a higher marginal cost) than would be the case in ideal circumstances (compare the ideal constraint in expression (10)). Given this new constraint, the "best nonideal society" — that is, the society with the highest normative value given this new realization constraint — is identified by optimizing the function

$$\mathcal{L}(e, f, s, \beta) = efs - \lambda\left(-1 + f + 2e^2 s^2\right). \tag{20}$$

33 Of course, in the real world, a shift from ideal to nonideal circumstances could lead to the imposition of all three kinds of constraint. We consider these different types of constraint in isolation for the purposes of analyzing their distinctive implications for normative reasoning.

This exercise yields the new first-order necessary conditions for the best nonideal society:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial e} &= fs - \lambda 4es^2 = 0 \\ \frac{\partial \mathcal{L}}{\partial f} &= es - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial s} &= fe - \lambda 4se^2 = 0\end{aligned} \tag{21}$$

Solving these equations implies three nonideal level principles: $e = s \approx 0.64$ and $f \approx 0.67$. Rearranging these first-order conditions implies two nonideal ratio principles:

$$\frac{f}{e} = 4es^2 \qquad\qquad \frac{f}{s} = 4se^2. \tag{22}$$

We now have a clear comparison between the ideal and nonideal prescriptions. Given a constraint that requires us to operate under "nonideal" opportunity costs, we should:

- maintain the ideal level of freedom but decrease (relative to the ideal) the levels of equality and security;
- increase (relative to the ideal) both the freedom-to-equality ratio and the freedom-to-security ratio.

Our nonideal prescriptions ostensibly deviate from our ideal principles. But recall that the question is not whether nonideal prescriptions deviate from ideal principles; some deviation is to be expected given the imposition of a constraint. The pertinent question is whether our nonideal prescriptions deviate from ideal principles *in surprising and unexpected ways*. I submit that the differences here are exactly as we would expect given the nature of the constraint. Our constraint leads to an increase in the relative costs of realizing equality and security; the result is a straightforward adjustment of equality-related and security-related principles — namely, that we should realize relatively less equality and security. To bolster the claim that these adjustments are exactly as we would expect, we note that our nonideal ratio principles satisfy the meta-principle noted at the end of section 3.2, reproduced here:

> The normatively ideal balance between two criteria x and y (i.e., the content of the ratio principle that governs x and y) is determined by the relative opportunity costs of their respective realization.

(See the appendix for the explanation why our nonideal ratio principles satisfy this meta-

principle.) Because the content of (e.g.) both our ideal and nonideal freedom-to-equality ratio principles is determined by the relative opportunity costs of realizing freedom and equality (similarly for the freedom-to-security principles), the differences between the ideal and nonideal ratio principles are exactly what we would expect in accordance with this meta-principle. Thus, shifting from "ideal" to "nonideal" opportunity costs raises no concerns about unexpected deviations from the ideal. Indeed, if this is the only constraint on our efforts to realize the ideal, then our nonideal principles prescribe the adjustments we would expect given our analysis of the ideal.

4.2. Constraining level principles. Suppose we model the transition from ideal to nonideal theory not as a change in the opportunity costs of realizing a particular criterion, but as a hard limit on the extent to which this criterion can be realized. For concreteness, assume that circumstances limit the extent to which we can realize equality, so that $e \leq 0.6$ (recall that, ideally, $e \approx 0.76$). Now we identify the best nonideal society by optimizing the amended function

$$\mathcal{L}(e, f, s, \beta, \mu) = efs - \beta(-1 + f + e^2 s^2) - \mu\left(e - \frac{6}{10}\right). \tag{23}$$

This yields the first-order necessary conditions for the best nonideal society:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial e} &= fs - \beta 2es^2 - \mu = 0 \\ \frac{\partial \mathcal{L}}{\partial f} &= es - \beta = 0 \\ \frac{\partial \mathcal{L}}{\partial s} &= fe - \beta 2se^2 = 0 \end{aligned} \tag{24}$$

Solving these implies three nonideal level principles: $e = 0.6$, $f \approx 0.67$, $s \approx 0.96$. Rearranging the first-order conditions implies two proportionality conditions:

$$\frac{f}{e} = 2es^2 + \frac{\mu}{\beta} \qquad\qquad \frac{f}{s} = 2se^2. \tag{25}$$

Again, compare these nonideal prescriptions with our ideal principles. If we are constrained to deviate from the ideal level of equality, we should:

- continue to realize the ideal level of freedom but increase (relative to the ideal) the level of security;
- maintain the ideal freedom-to-security ratio while increasing (relative to the

ideal) the freedom-to-equality ratio.

Again, our nonideal prescriptions deviate from our ideal prescriptions. But, again, the pertinent question is whether these deviations are *unexpected in light of the stipulated constraint*. And I submit that these differences are entirely as expected. First, with respect to the differences in level principles: Given our specified normative value function F, we should expect that, if we are constrained to realize less equality than we would in the ideal society, then a nonideal society will need to increase the realization of at least one of the remaining criteria to compensate for the decrease in equality. So the increase of security is to be expected. Second, with respect to the differences in ratio principles: Notice that only the nonideal freedom-to-equality ratio prescription deviates from the ideal; the nonideal freedom-to-security ratio prescription satisfies the ideal principle. This contrasts with the case where we are constrained to leave one of our ideal ratio principles unsatisfied (section 3.3). In that case, our nonideal principles prescribe that we leave *all* ratio principles unsatisfied, even those that are unconstrained. Yet in this case, where we are constrained to leave one of our ideal level principles unsatisfied, our nonideal principles prescribe that we should maintain the ideal ratio of the unconstrained criteria (viz., freedom and security). This is as we might expect: when we are constrained to leave an ideal level principle unsatisfied, we would expect to adjust the balance between this criterion and the other criteria while nonetheless leaving the ideal balance of the unconstrained criteria untouched. Thus, our conclusion here is similar to that in section 4.1: if the only constraint on our efforts to realize the ideal is a constraint on an ideal level principle, then our nonideal principles prescribe the adjustments we would expect given our analysis of the ideal.

4.3. Summary. The lesson of the analyses in the previous two subsections is that the second best theorem's warning about surprising and unexpected deviations from the ideal is triggered by a specific type of constraint on realizing the ideal.[34] This stands in sharp contrast with political theorists' frequent suggestions that the second best theorem warns of unexpected deviations given any type of "nonideal" constraint. While we can extend the reach of the second best theorem beyond applied modes of normative reasoning, we must be aware that its significance is also more specific than political theorists have typically realized. In particular, the warning about unexpected deviations from the ideal so often associated with the theorem only arises in the presence of, specifically, a constraint on

34 Wiens's (2016) criticisms of the "background assumptions" interpretation of the theorem gesture at this point, although he is not precise enough about distinguishing between different kinds of constraints on realizing ideal principles.

realizing the ideal balance of two normative criteria. Second best normative reasoning in the presence of other types of constraints, far from triggering concerns about unexpected deviations from the ideal, generate nonideal prescriptions that are entirely expected given the nature of the constraint and our analysis of the ideal society.

5. CODA

This paper contains an unusually high number of mathematical expressions for a work of political philosophy. Some might object that "all this math" is unnecessary, but in fact it is unavoidable. My main purpose is to show how the theory of second best can be extended to new domains simply by developing new interpretations of the mathematical model Lipsey and Lancaster use to prove their theorem. We need to see this model in action if we are to develop intuitions about plausible interpretations — indeed, if we are to even recognize the possibility of non-standard interpretations at all. Scholars have generally failed to recognize alternative interpretations of the theorem because they have focused exclusively on a specific interpretation of the model. For economists, the mathematics is so familiar that they move without pause between the model and its economic interpretation, whereas the mathematics is so unfamiliar for political theorists that they can't but rely on economists' informal glosses on the theorem and its implications. Whatever the case may be, we are blinded to the full generality of the theory of second best if we focus exclusively on its most familiar interpretation.

To develop an ideal-theoretic interpretation of the theory of second best, I put the mathematical model into action twice: once in section 2 to pry apart the mathematics from its economic interpretation, and a second time in section 3 to make plausible the thought that the mathematics can be interpreted using concepts from ideal normative theory. Yet, for all the mathematical expressions, it is crucial to understand that this paper does very little mathematical work. This is because the mathematics on its own implies *nothing* about ideal normative theory (nor about economic theory for that matter). In deriving normative theoretical insights from the model, all the action is in how we interpret the objects in the model. Accordingly, my attention throughout the paper is entirely focused on conceptual issues.

None of this is to say that the mathematics is idle.[35] To begin with, the mathematics enables us to give precise definitions to abstract concepts, such as the *overall normative value* of possible social arrangements, or the *opportunity costs* involved in realizing

35 Much of what I have to say in this paragraph is indebted to, among others, Ingham (2015); Rodrik (2015); Rubinstein (2012).

normative criteria, or the *ideal balance* among normative criteria. This isn't to suggest that our first-order normative reasoning about ideal societies can or even should attain mathematical precision. It is instead a means to climbing up one level of abstraction so that we can get a higher-order view on our first-order normative reasoning and thereby investigate some of general features of that first-order reasoning. The abstract concepts we use to theorize more generally about our first-order normative reasoning need to cover a diverse range of particular instances. To wit, the concept of an *overall normative value function* covers multifarious ways of comparatively evaluating possible societies, which might appear on their surfaces to have very little in common. By defining this concept precisely in terms of its core logical attributes (piggy-backing on the logical attributes of the function we use to represent the concept), we can say with some precision which particular instances of first-order normative reasoning are covered by the model and which are not (see section 4). In addition to precise definitions, the mathematics disciplines our (second-order) reasoning about the implications of setting these concepts in certain relationships to each other, as happens when we use a particular mode of (first-order) reasoning to specify the defining attributes of ideal societies. This discipline enables us to uncover significant insights that would otherwise be hidden from view. For example, once one absorbs the basic anti-approximation lesson of the theory of second best, it is intuitive to think that any constraint on realizing the ideal triggers a warning against approximation. The mathematics enables us (again, in section 4) to precisely define different types of constraints on realizing the ideal and show, contrary to intuition, that some familiar constraints do not in fact trigger a warning against approximation — that in fact all but a very specific and underappreciated type of constraint generate nonideal principles that prescribe deviations from the ideal that are exactly as we would expect.[36]

The mathematics is not an end in itself. It is a means to articulating and exposing conceptual insights that would otherwise be unavailable, or at least significantly more difficult to grasp.

6. APPENDIX

Our model in section 3 presents two ways to characterize an ideal society in terms of general normative principles. A third, more abstract kind of characterization is important for showing how our model of normative reasoning exemplifies Lipsey and Lancaster's more general model. This more abstract property can be exposed by attending to a more

36 For an additional illustration of how the mathematics enables us to uncover otherwise obscured insights, see my remarks at the beginning of section 2.

general description of the conditions expressed in (12) and (13). We start by noticing that the conditions expressed in (12) can be rewritten as

$$\begin{aligned} F_e - \lambda G_e &= 0 \\ F_f - \lambda G_f &= 0 \\ F_s - \lambda G_s &= 0 \end{aligned} \tag{26}$$

where F_i is the partial derivative of F with respect to $i = e, f, s$ and G_i is the partial derivative of G with respect to i (compare these conditions with those in (2) above). Note that $F_e = fs$, $F_f = es$, $F_s = ef$, $G_e = 2es^2$, $G_f = 1$, and $G_s = 2se^2$. These conditions imply that the equalities in (13) can be rewritten as

$$\frac{F_s}{F_f} = \frac{G_s}{G_f} \qquad\qquad \frac{F_e}{F_f} = \frac{G_e}{G_f} \tag{27}$$

(compare with (3) above). These expressions reveal an important "second-order attribute" of the ideal society, namely, that the relative significance assigned to our three criteria by the ratio principles in (13) is determined by certain marginal conditions that indicate the manner in which these criteria contribute to overall normative value (as encoded by F) and the trade-offs involved in realizing them (as encoded by G). As a first pass, we can say that the ideal society satisfies the following *meta-principle*:

> The normatively ideal balance between two criteria x and y (i.e., the content of the ratio principle that governs x and y) is determined by the relative opportunity costs of their respective realization.

To gain a more precise understanding of this principle and its significance, we need a substantive interpretation for the partial derivatives of F and G. I do this using an example that makes clear how the interpretation applies more widely.

The partial derivative of F with respect to the variable f, denoted F_f, indicates the marginal contribution that freedom makes to our estimation of a society's overall normative value as encoded by the value function F. To start simple, if $F_f = f$, this means that an arbitrarily small increase (decrease) in a society's level of freedom produces an increase (decrease) in that society's overall normative value that is equal to the amount of the increase times the status quo level of freedom, assuming that we hold fixed its level of equality and security. In our example above, $F_f = es$, which means that an arbitrarily small increase (decrease) in a society's level of freedom produces an increase (decrease) in that society's overall normative value that is equal to the size of the increase times

the product of the status quo levels of equality and security, assuming that we hold the latter constant. Concretely, if the current levels of equality and security are 0.6 and 0.25 respectively, then $F_f = es$ implies that a small increase in the level of freedom equal to $\epsilon > 0$ produces a $0.6 \times 0.25 \times \epsilon = 0.15\epsilon$ increase in overall normative value. Given a similar understanding of F_e and F_s as indicating the marginal value contributions of equality and security respectively, an expression such as $\frac{F_s}{F_f}$ indicates a ratio of two variables' marginal contributions to overall normative value. We can interpret this ratio as the rate at which we can exchange a small amount of security for a small amount of freedom (or vice versa) without changing a society's overall normative value. In our example, $\frac{F_s}{F_f} = \frac{ef}{es} = \frac{f}{s}$, which implies the following: If the current level of freedom is 0.4 and the current level of security is 0.2, then we can exchange arbitrarily small amounts of security for freedom at a rate of 2 to 1 without changing society's overall normative value. Roughly, if we decrease (increase) security by 0.02, then we must increase (decrease) freedom by 0.01 to maintain the current overall normative value. In sum, the expression $\frac{F_s}{F_f}$ indicates the (marginal) exchange rate between security and freedom that preserves a society's overall normative value. Call this the *value-preserving exchange rate* for convenience.

The partial derivative of G with respect to the variable s, denoted G_s, indicates the marginal opportunity cost of a small increase of security, where this cost is encoded by the realization function G and expressed in terms of an implied loss of equality and security. Given a similar understanding of G_e and G_f as indicating the marginal opportunity cost of realizing equality and freedom respectively, an expression such as $\frac{G_s}{G_f}$ indicates the relative opportunity costs of realizing more security versus more freedom. In our example, $\frac{G_s}{G_f} = \frac{2se^2}{1}$, which implies the following: If we redirect resources from the realization of freedom to the realization of security, then, to satisfy the constraint $G = 0$, increasing security by 0.01 will require us to decrease freedom by $2se^2 \times 0.01 = 0.02se^2$, with s and e being the status quo levels of security and equality (and assuming that we hold these fixed). In sum, the expression $\frac{G_s}{G_f}$ indicates the (marginal) rate at which a society can redirect resources from the realization of security to the realization of freedom while continuing to satisfy the realization constraint. Call this the *constraint-preserving transformation rate* for convenience.

Given these interpretations, the expressions in (27) say that the ideal society is such that the value-preserving exchange rate for security and freedom is equal to the constraint-preserving transformation rate for security and freedom, and similarly for equality and freedom. Now recall two observations: first, that we can interpret the expressions in (13) as ratio principles, which indicate the ideal balance of two criteria; second, that the expressions in (27) are simply a more general way to express the ratio principles

in (13). It follows that the ideal society is characterized by the second-order attribute that the content of ideal ratio principles is determined by value-preserving exchange rates and constraint-preserving transformation rates. In other words, the ideal society is characterized by the meta-principle stated above.

Let's now turn to several observations about what happens under various types of constraints on realizing ideal principles.

- The analysis in section 3.3 assumes a constraint on realizing an ideal ratio principle. Given this constraint, the second best society not only fails to satisfy the ideal ratio principles in (13), but it also fails to satisfy the meta-principle across the board (with respect to every pair of variables). This can be seen by comparing (17) with (27). Thus, a constraint on realizing an ideal ratio principle requires the second best prescriptions to deviate from the ideal prescriptions in a deep and pervasive way.

- The analysis in section 4.1 assumes a constraint on the realization function, which represents a shift to the opportunity costs that arise in "nonideal" circumstances. Given this constraint, the second best society violates the ideal freedom-to-security ratio principle. Yet the second best society satisfies the meta-principle with respect to all pairs of criteria. This can be readily seen once we note that the partial derivatives of G with respect to e and s are (respectively) $4es^2$ and $4se^2$. Thus,

$$\frac{F_e}{F_f} = \frac{fs}{es} = \frac{4es^2}{1} = \frac{G_e}{G_f} \qquad\qquad \frac{F_s}{F_f} = \frac{ef}{es} = \frac{4se^2}{1} = \frac{G_s}{G_f},$$

  both of which satisfy the meta-principle (compare with (22)). Thus, in this case, the second best prescriptions deviate from the ideal prescriptions in a way that is ultimately constrained by the meta-principle and, thus, as we should expect.

- The analysis in section 4.2 assumes a constraint on realizing the ideal equality level principle. Given this second best constraint, the second best society violates the ideal freedom-to-equality ratio principle. It also violates the meta-principle with respect to freedom and equality:

$$\frac{F_e}{F_f} = \frac{G_e + \frac{\mu}{\beta}}{G_f} \neq \frac{G_e}{G_f}$$

  (compare with (25)). Yet the second best society satisfies the meta-principle with

respect to the unconstrained criteria freedom and security:

$$\frac{F_s}{F_f} = \frac{ef}{es} = \frac{2se^2}{1} = \frac{G_e}{G_f}.$$

Thus, a constraint on realizing an ideal level principle requires that the second best prescriptions violate the meta-principle with respect to the constrained criterion but not at all with respect to the unconstrained criteria.

REFERENCES

Aristotle. 1988. *The Politics*. Cambridge University Press.

Blackorby, Charles, Russell Davidson and William Schworm. 1991. "The Validity of Piecemeal Second-Best Policy." *Journal of Public Economics* 46:267–290.

Boadway, Robin. 2017. "Second-Best Theory: Ageing Well at Sixty." *Pacific Economic Review* 22(2):249–70.

Brennan, Geoffrey and Philip Pettit. 2005. The Feasibility Issue. In *Oxford Handbook of Contemporary Philosophy*, ed. Frank Jackson and Michael Smith. Oxford: Oxford University Press.

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

Cohen, G.A. 2003. "Facts and Principles." *Philosophy & Public Affairs* 31(3):211–245.

Coram, Bruce Talbot. 1996. Second Best Theories and the Implications for Institutional Design. In *The Theory of Institutional Design*, ed. Robert E. Goodin. New York: Cambridge University Press.

Davis, Otto A. and Andrew B. Whinston. 1965. "Welfare Economics and the Theory of Second Best." *The Review of Economics and Statistics* 32(4):1–14.

Davis, Otto A. and Andrew B. Whinston. 1967. "Piecemeal Policy in the Theory of Second Best." *Review of Economics and Statistics* 34(3):323–31.

Debreu, Gerard. 1959. *Theory of Value*. New York: Wiley.

Estlund, David. 2011. "Human Nature and the Limits (If Any) of Political Philosophy." *Philosophy & Public Affairs* 39(3):207–237.

Estlund, David. 2018. "Utopophobia: On the Limits (If Any) of Political Philosophy." unpublished manuscript, Brown University.

Gilabert, Pablo. 2012. "Comparative Assessments of Justice, Political Feasibility, and Ideal Theory." *Ethical Theory & Moral Practice* 15(1):39–56.

Goodin, Robert E. 1995. "Political Ideals and Political Practice." *British Journal of Political Science* 25(1):37–56.

Heath, Joseph. 2013. "Ideal Theory in an Nth-Best World: The Case of Pauper Labor." *Journal of Global Ethics* 9(2):159–72.

Hoff, Karla. 2000. Second and Third Best Theories. In *Reader's Guide to the Social Sciences*, ed. Jonathan Michie. Vol. 1 New York: Routledge pp. 1463–64.

Ingham, Sean. 2015. "Theorems and Models in Political Theory: An Application to Pettit on Popular Control." *The Good Society* 24(1):98–117.

Ingham, Sean. 2019. "Why Arrow's Theorem Matters For Political Theory — Even If Preference Cycles Never Occur." *Public Choice* 179(1):97–111.

Jewitt, Ian. 1981. "Preference Structure and Piecemeal Second Best Policy." *Journal of Public Economics* 16:215–231.

Lipsey, R. G. and Kelvin Lancaster. 1997. The General Theory of Second Best. In *The Selected Essays of Richard G. Lipsey, Volume 1: Microeconomics, Growth and Political Economy*, ed. Richard G. Lipsey. Cheltenham: Edward Elger.

Lipsey, R.G. and Kelvin Lancaster. 1956. "The General Theory of Second Best." *The Review of Economic Studies* 24(1):11–32.

Lipsey, Richard G. 2007. "Reflections on the General Theory of Second Best at its Golden Jubilee." *International Tax and Public Finance* 14:349–364.

Ng, Yew-Kwang. 2004. *Welfare Economics: Towards a More Complete Analysis*. New York: Palgrave Macmillan.

Räikkä, Juha. 2000. "The Problem of the Second Best: Conceptual Issues." *Utilitas* 12(2):204–218.

Rawls, John. 1999. *A Theory of Justice*. 2 ed. Cambridge, MA: Harvard University Press.

Rodrik, Dani. 2015. *Economics Rules: Why Economics Works, When It Fails, and How To Tell The Difference*. Oxford: Oxford University Press.

Rossi, Enzo and Matt Sleat. 2014. "Realism in Normative Political Theory." *Philosophy Compass* 9(10):689–701.

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

Sangiovanni, Andrea. 2016. "How Practices Matter." *Journal of Political Philosophy* 24(1):3–24.

Sen, Amartya. 2009. *The Idea of Justice*. Cambridge, MA: Harvard University Press.

Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy & Public Affairs* 38(1):5–36.

Southwood, Nicholas. forthcoming. "The Feasibility Issue." *Philosophy Compass* .

Stemplowska, Zofia and Adam Swift. 2012. Ideal and Nonideal Theory. In *The Oxford Handbook of Political Philosophy*, ed. David Estlund. New York: Oxford University Press.

Swift, Adam. 2008. "The Value of Philosophy in Nonideal Circumstances." *Social Theory and Practice* 34(3):363–387.

Tessman, Lisa. 2010. "Idealizing Morality." *Hypatia* 25(4):797–824.

Valentini, Laura. 2012. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7(9):654–664.

Valentini, Laura. 2017. On the Messy 'Utopophobia vs. Factophobia' Controversy: A Systematization and Assessment. In *Political Utopias: Contemporary Debates*, ed. Kevin Vallier and Michael Weber. New York: Oxford University Press pp. 11–35.

Wiens, David. 2016. "Assessing Ideal Theories: Lessons from the Theory of Second Best." *Politics, Philosophy and Economics* 15(2):132–149.