

## IDEAL THEORY AND THE THEORY OF SECOND BEST

---

David Wiens

**Abstract.** Philosophers occasionally invoke Lipsey and Lancaster's "general theory of second best" to challenge the *ideal guidance view*, the view that ideal political principles can provide normative guidelines for our efforts to address injustice amidst unfavorable circumstances. Roughly, the theorem says: if certain conditions are met, then what we should do in nonideal circumstances does not necessarily approximate what we should do in ideal circumstances. But extant challenges to the ideal guidance view are based on mistaken interpretations of the theorem's antecedent condition. I show that, once we understand the antecedent condition correctly, the theory of second best does not present as tough a challenge to the ideal guidance view as is typically believed.

### 1. INTRODUCTION

Philosophers occasionally invoke Lipsey and Lancaster's "general theory of second best"<sup>1</sup> to challenge the *ideal guidance view*, the view that ideal political principles can provide normative guidelines for our efforts to address injustice amidst unfavorable circumstances.<sup>2</sup> Roughly, the theorem says: if certain conditions are met, then what we should do in nonideal circumstances does not necessarily approximate what we should do in ideal circumstances.<sup>3</sup> But extant challenges to the ideal guidance view are based on mistaken interpretations of the theorem's antecedent condition.<sup>4</sup> This matters because

1 Lipsey and Lancaster (1956) (henceforth L&L) present the original statement of the general theory of second best. Ng (2004, ch. 9) provides an accessible introduction.

2 For debate on the ideal guidance view, see, among others, Farrelly (2007); Mason (2004); Sen (2009); Simmons (2010); Valentini (2009); Wiens (2012).

3 I'll use "theorem" to refer to the centerpiece of L&L, viz., the theorem proved in section 7 of that paper; I'll use "theory" to refer to the entity composed of the theorem, its proof, and the implications drawn therefrom.

4 Some philosophers have also expressed confusion about the modal strength of the theorem's consequent. On the "necessarily-not" reading, the theorem says: if the antecedent condition is satisfied, then, necessarily, a second best outcome does not satisfy the conditions required for a first best outcome. On the "not-necessarily" reading, the theorem says: if the antecedent condition is satisfied, then a second best outcome does not necessarily satisfy the conditions required for a first best outcome. Philosophers have typically interpreted L&L as endorsing the necessarily-not version, but have hesitated to apply anything stronger than the not-necessarily version to political theory. Since I think L&L prove only the not-necessarily version of the theorem, and since philosophers' challenge to the ideal guidance view depends only on the not-necessarily version, I leave this issue aside.

the theorem's implications for normative political philosophy depend on the nature of cases to which it applies, as specified by its antecedent conditions. I show that, once we understand the antecedent condition correctly, the theory of second best does not present as tough a challenge to the ideal guidance view as is typically believed.

The paper proceeds as follows. Section 2 surveys extant applications of the theory of second best to normative political theory and outlines what's at stake for political philosophers in interpreting the theorem correctly. Section 3 introduces the notion of an optimization problem and explores at some length the extent to which familiar exercises in normative political philosophy are isomorphic to the sort of optimization exercise presupposed by the theorem's proof, something that has yet to be shown. Section 4 discusses the theorem's antecedent in more detail, while section 5 elaborates the application of second best reasoning to normative political philosophy. Section 6 concludes by drawing out the implications of L&L's theorem for philosophical methodology, arguing that skeptics would do best to abandon appeals to the theory of second best in their efforts to deny the ideal guidance view.

## 2. THE THEORY OF SECOND BEST IN NORMATIVE POLITICAL THEORY

As L&L originally gloss it,

The general theorem of the second best states that if one of the Paretian optimum conditions cannot be fulfilled a second best optimum situation is achieved only by departing from all other optimum conditions.<sup>5</sup>

At face value, the theorem bears on our investigation of the economic policies we should adopt when we are prevented from satisfying the conditions required for a Pareto efficient outcome. In fact, the theorem is supposed to apply more generally, to "all maximization problems". But just how it applies to political theory is not obvious. In general, political theorists deploy the theorem of second best to challenge the *ideal guidance view*, which can be reduced to the slogan, "ideal theory provides guidance for nonideal theory".<sup>6</sup> Here, *ideal theory* is roughly conceived as specifying the requirements for a fully just scheme of social and political institutions and *nonideal theory* is understood to specify how we

<sup>5</sup> L&L, 12. Clearly, L&L's language here suggests what I've called (in footnote 4) the necessarily-not version of the theorem. Elsewhere, their gloss on the consequent is more ambiguous: "the other Paretian conditions, although still attainable, are, in general, no longer desirable" (L&L, 11). I mention this issue only to set it aside. I follow economists and other philosophers in deploying the not-necessarily version of the theorem throughout.

<sup>6</sup> Cf. Wiens (2012).

## THE THEORY OF SECOND BEST

should respond to injustice amidst nonideal circumstances. Philosophers' deployment of the theorem against this view follows a general schema:

- (1) *The Theorem of Second Best*. If certain conditions can't be satisfied, then we should not necessarily apply ideal normative principles in nonideal circumstances.
- (2) The relevant conditions can't be satisfied.
- (3) Therefore, we should not necessarily apply ideal principles in nonideal circumstances.

What this argument implies for the ideal guidance view depends on how we fill in this schema, in particular, how we interpret the theorem's antecedent.

There are three main ways in which philosophers have made this schema more determinate. One argument deploys a *theoretical presuppositions* interpretation of the theorem's antecedent. This says that the theorem's consequent applies when the conditions assumed in specifying ideal normative principles do not obtain: "*If any one of the conditions presupposed by ideal theory is missing, then the Theory of Second-Best warns that we might... need to make systematic alterations right across the board in the prescriptions of ideal theory*".<sup>7</sup> Typically, ideal theory assumes conditions that are unlikely to obtain in the actual world; for example, when specifying his principles of justice, Rawls notably assumes that individuals will generally comply with the selected principles, that society has sufficient material resources to fulfill the principles' requirements, and that society is self-contained (i.e., there are no cross-border transactions).<sup>8</sup> Since the conditions presupposed when specifying ideal principles typically do not obtain in the actual world, the theoretical presuppositions interpretation implies that we should not necessarily apply ideal principles in nonideal circumstances.

A second argument deploys a *moral ideals* interpretation, which says that the theorem's consequent applies when we cannot realize some of our moral and social ideals: "*When our ideals cannot all be realized simultaneously, the general theory of the second best... warns us against assuming naively that it is better to implement more of our ideals rather than fewer (or indeed to implement each of them to a greater rather than lesser degree)*".<sup>9</sup> It is almost certainly true that, given our actual circumstances, we cannot satisfy all of our ideals simultaneously (e.g., liberty, equality, fraternity, and material prosperity, to take Goodin's example). It thus follows from the moral ideals interpretation that we

<sup>7</sup> Goodin (2012, 162, emphasis added).

<sup>8</sup> Rawls (1999, 7, 8).

<sup>9</sup> Goodin (1995, 54, emphasis added).

should not necessarily seek to satisfy as many of our ideals as we can or satisfy them to the greatest extent possible.

A third argument deploys an *ideal institutions* interpretation, which says that the theorem's consequent applies when we are unable to (fully) implement the institutional scheme that is appropriate for ideal conditions:

[W]hat the theory of the second best suggests is that... [s]ince *departure from any one condition in the institution used for the model* [of the best scheme] means that all the other conditions may not be desirable, it is not clear whether the optimum choice is to get as close to the original as possible, or to construct a completely different institution.<sup>10</sup>

Given that it is generally infeasible to realize an ideal institutional scheme, the ideal institutions interpretation implies that we should not necessarily try to implement an ideal institutional scheme in nonideal circumstances.

The preceding shows that distinct interpretations of the theorem's antecedent yield distinct implications for the ideal guidance view. Although these implications are clearly related, their differences are important because they provoke distinct responses from a theorist who insists that ideal principles provide useful guidance for nonideal theory. For example, to the moral ideals interpretation, the ideal guidance proponent concedes that we cannot simultaneously satisfy our ideals but replies that ideal theory is motivated in part by this recognition. This is why ideal theory is supposed to help nonideal theory—ideal theory delivers principles that putatively help us determine the relative weight we should give to disparate ideals.<sup>11</sup> To the ideal institutions interpretation, the ideal guidance proponent replies that we should not try to implement ideal *institutions* in nonideal circumstances; rather, we should try to implement the institutional scheme that best satisfies the requirements of our ideal *principles*. There is no inconsistency here because ideal normative principles do not have any determinate institutional implica-

10 Coram (1996, 93, emphasis added). Coram goes on to claim that the theory of second best warns us of two fallacies: “the fallacy of continuity”, which holds that similar initial conditions produce similar results; and “the fallacy of stretchability”, which says that small changes to institutions leads to small changes in the outcome (Coram, 1996, 94). These are surely fallacies, as Coram's examples show; but they apply in ideal circumstances as much as they apply in nonideal circumstances. To wit, assuming we can fulfill the conditions required to achieve Pareto efficiency, small changes to the initial economic endowment can produce very different Pareto efficient outcomes. Hence, the insight of the theory of second best is not, *pace* Coram, that “radical alterations in institutions may be required to accommodate small shifts in initial conditions” (Coram, 1996, 91). That insight bears on the first best case too.

11 See, e.g., Gilabert (2012); Swift (2008).

tions.<sup>12</sup> Interpreted in either of these ways, the theorem of second best poses little threat to the ideal guidance view.

The argument based on the theoretical presuppositions interpretation poses a greater threat to the ideal guidance view. On this interpretation, ideal principles satisfy the theorem's antecedent by definition; however one defines "ideal theory", it is true that the specification of ideal normative principles abstracts from certain nonideal features of the actual world, (implicitly) assuming circumstances that are unlikely to obtain.<sup>13</sup> Indeed, if our specification of ideal principles assumed circumstances that sufficiently resembled the salient features of our decidedly nonideal world, it would be difficult to grasp what is meant by "ideal theory". Given this, the theoretical presuppositions interpretation implies that we should not necessarily adopt ideal principles as normative guidelines for addressing injustice in a nonideal world. Since we can only determine whether ideal principles are appropriate guidelines upon assessing the likely consequences of implementing them in our actual circumstances, ideal principles as such provide unreliable guidance and are, thus, useless guides. If the theoretical presuppositions interpretation is correct, the ideal guidance proponent has no recourse except to argue that the theorem of second best does not apply to normative political theory, that it is not the type of maximization exercise to which the theorem applies. In the next section, I show that familiar forms of political theory can be naturally cast as the right sort of maximization problem before returning in section 4 to discuss whether the theoretical presuppositions interpretation of the theorem's antecedent is correct.

### 3. IDEAL THEORY AS AN OPTIMIZATION EXERCISE

L&L claim that their theorem is general enough to apply to "all maximization problems, not just welfare theory".<sup>14</sup> But if L&L's theorem is to threaten the ideal guidance view, we must show in particular that familiar practices for specifying ideal normative principles are isomorphic to the sort of maximization problem to which the theorem applies.

A constrained optimization problem has four core components: an *objective*, a specification of the outcome of ultimate pursuit; a set  $X$  of *choice variables*; an *objective function*, which specifies how the objective depends on the elements of  $X$ ; and a *constraint function*, which defines the set of feasible outcomes by specifying the set of values that can be assigned to the elements of  $X$ . The problem of consumer choice provides

<sup>12</sup> See, e.g., Valentini (2011).

<sup>13</sup> Cf. Hamlin and Stemplowska (2012); Valentini (2012).

<sup>14</sup> L&L, 12, n. 2.

a simple illustration. Suppose you are faced with a choice of two goods, so  $X = \{x, y\}$ . Ultimately, you don't care about the amount of each good you consume, but the level of utility you receive from consuming those goods. Your objective is to choose amounts of  $x$  and  $y$  that maximize your utility; your objective (utility) function specifies how your utility depends on  $x$  and  $y$ . Of course, like every consumer, you can only consume as much  $x$  and  $y$  as your budget allows given the unit prices for  $x$  and  $y$ ,  $p_x$  and  $p_y$  respectively. Let your utility function be  $u(x, y) = 2xy$ . Suppose you have \$100 to spend and  $p_x = \$5$  and  $p_y = \$10$ ; then your constraint function is  $5x + 10y \leq 100$ . Solving this, we derive necessary conditions for maximizing utility under the given constraints: you must buy 10 units of  $x$  and 5 units of  $y$ .

To further illustrate the structure of an optimization problem, take a more complicated example: the theory of Pareto optimality, L&L's central concern. Welfare economists start with the question "What conditions must hold to realize an efficient allocation of goods?" They stipulate that the answer is delivered by some function of individuals' utility, the inputs used to produce the goods and services that contribute to individuals' utility, and certain "natural" production constraints. Hence, the set of choice variables comprises the goods and services that circulate among economic actors; consumers choose how much of each good to consume and producers choose how much of each good to produce. But economists are not ultimately interested in the chosen levels of consumption and production per se. They are interested in characterizing a social state that exhibits the property of efficiency, defined as follows. Let  $X$  and  $X'$  denote two allocations of goods among the  $I$  individuals in the economy;  $u_i(x_i)$  and  $u_i(x'_i)$  denote the utility individual  $i$  receives from her bundle of goods given  $X$  and  $X'$  respectively.  $X$  is an *efficient* or *Pareto optimal* allocation if there is no alternative  $X'$  that satisfies both of the following conditions: (1)  $u_i(x'_i) \geq u_i(x_i)$  for all  $i = 1, 2, \dots, I$  (2)  $u_i(x'_i) > u_i(x_i)$  for at least one  $i$ . We can see, then, that the objective — Pareto optimality — is a function of individuals' utility functions and the goods they consume. Individuals' consumption possibilities are constrained by their budgets, firms' production functions and any limits on the availability of production inputs. (Notably, production and consumption externalities, asymmetric information, monopolies, and transaction costs are excluded from the set of assumed constraints.) Given these functions, we can derive the conditions that must hold if we are to attain a Pareto optimum, stated in terms of certain relationships that must obtain between the various goods that individuals consume, the prices of goods, and the inputs used to produce those goods.

*Paretian Optimum Conditions.* (1) For each pair of goods,  $x, y$ , the marginal rate at which individuals are willing to substitute  $x$  for  $y$  (the marginal rate of sub-

## THE THEORY OF SECOND BEST

stitution) must be the same for all individuals who consume  $x$  and  $y$ . (2) The marginal rate at which one production input can be substituted for another in production must be the same for all goods produced in the economy. (3) For each pair of goods,  $x$ ,  $y$ , consumers' marginal rate of substitution must equal the marginal rate of transformation, i.e., the rate at which the economy can redirect production of  $x$  into production of  $y$ .<sup>15</sup>

In sum, the theory of Pareto optimality has the following structural features: an objective (Pareto optimality); a set of choice variables (economic goods); a function that specifies how achievement of the objective depends on the choice variables (a social welfare function composed of individuals' utility functions); and a function that specifies the set of feasible outcomes (a social production function composed of firms' production functions and constraints on input availability). The goal of the theoretical exercise is to derive a set of *optimum conditions*, which specify the conditions that must obtain if we are to achieve our objective given the circumstances.

I now show that Rawls's (1999) theory shares these structural features and can, thus, be naturally understood as the sort of optimization exercise to which the theory of second best applies. Rawls's aim is to "examine the principles of justice that would regulate a well-ordered society".<sup>16</sup> Rawls defines a well-ordered society as "a system of cooperation designed to advance the good of those taking part in it"<sup>17</sup> Further, "it is a society in which (1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles".<sup>18</sup> Rawls limits his attention to identifying the constitutive features of the "basic structure" of a well-ordered society, which consists of those institutions that are responsible for defining individuals' rights and privileges and allocating "the benefits and burdens of social cooperation".<sup>19</sup> From all this, we can see that Rawls's theoretical enterprise takes the basic structure of a well-ordered society as a regulative objective, much in the way that economists take a Pareto optimal allocation as a regulative objective.

To aid his theoretical exercise, Rawls reflects on the social and moral ideals that are manifest by a fully just basic structure. Without any claim to be exhaustive, and in no particular order, Rawls describes the basic structure of a well-ordered society as one that: is constituted by *impartial* rules that are *acceptable* to rational individuals;

<sup>15</sup> See Ng (2004, ch. 2) for an accessible derivation of these conditions.

<sup>16</sup> Rawls (1999, 8).

<sup>17</sup> Rawls (1999, 4).

<sup>18</sup> Rawls (1999, 4).

<sup>19</sup> Rawls (1999, 4).

engenders outcomes that *harmonize diverse interests*; respects individuals' status as *equals*; and yields a *fair division* of gains from cooperation.<sup>20</sup> On Rawls's view, a society's institutional scheme is just to the extent that it realizes these fundamental ideals; that is, the moral value of an institutional scheme depends on the extent to which it manifests impartiality, rational acceptability, respect for individuals, social equality, and the like. These fundamental moral and social ideals compose the set of choice variables.

This might seem an odd characterization of Rawls's view, since he explicitly has parties to the social contract compare and choose among principles for governing institutional arrangements.<sup>21</sup> But the point of this comparative exercise is determining the extent to which the fundamental moral and social ideals he adopts are realized by the various institutional alternatives that arise from implementing distinct normative principles. In choosing a set of principles and, hence, the range of institutional schemes that satisfy those principles, we are choosing the extent to which the resultant social states of affairs realize our fundamental ideals. It is true that Rawls does not explicitly treat his fundamental ideals as choice variables and he does not explicitly search for the optimal degree of realization for each. Instead, he begins by stipulating an objective, defined as an institutional scheme with certain properties (those that constitute a well-ordered society, defined in terms of his fundamental ideals). He then constructs a hypothetical choice situation (the original position) to model the optimal realization of his fundamental ideals and then argues that his principles of justice are (maximally) consistent with his model. But, in all this, Rawls proceeds exactly as the economist does. The latter stipulates an objective, defined as the realization of a certain property (efficiency, defined in terms of individuals' utility functions) and then models a situation that realizes this property to derive the conditions that are consistent with this model.

In principle, Rawls's fundamental ideals can be realized by degrees. For example, we can model greater or lesser impartiality by altering the information available behind the veil of ignorance; we can model greater or lesser respect for individuals' equality by altering the weights given to individuals or subgroups in collective decision-making; we can model greater or lesser rational acceptability by altering the range of considerations that form the basis for individuals' choices.<sup>22</sup> As we adjust the extent to which Rawls's fundamental ideals are manifest in the model, we almost certainly alter the principles that would be chosen.<sup>23</sup> For example, if we alter the rational basis for individuals' choices

20 Rawls (1999, secs. 1–4).

21 Rawls (1999, sec. 21).

22 Rawls (1999, 126f) acknowledges this point when discussing alternative constructions of the original position.

23 Cf. Rawls (1999, 105).



by including the probability distribution across social positions individuals may occupy, we'd likely arrive at the principle of average utility.<sup>24</sup> Or, if we permitted people to know which social position they are likely to occupy and grant those who are likely to be well-off greater weight in collective decision-making, we'd likely arrive at aristocratic rather than democratic principles. My point here is twofold: first, that Rawls's fundamental ideals can be realized by degrees; and, second, that different sets of principles realize Rawls's fundamental ideals to differing extents. The extent to which Rawls's objective is realized is a function of the extent to which his fundamental ideals are realized. We may appear to be choosing a set of principles in the original position; but, since normative principles encode our fundamental ideals in different ways, we are ultimately choosing the degree to which our fundamental ideals are realized. This supports my contention that Rawls's fundamental ideals compose the set of choice variables.

As I've already noted, Rawls's objective is a function of certain fundamental ideals; the extent to which a society's institutional scheme manifests justice depends on the extent to which his fundamental ideals are manifest. This function should assign weights to the constituent ideals, which encode each ideal's relative contribution to the moral value of an institutional scheme, as well as the rate at which we should be willing to trade greater realization of one ideal for lesser realization of another. The objective function should also specify any functional interdependencies that exist among the constituent ideals. In particular, we need to specify how (if at all) the contribution made by one ideal to the value of an institutional scheme depends on the realization of other ideals. For example, the value to an individual or society of individual liberty might depend on the extent to which other ideals (e.g., equality, security, community, etc.) are manifest. Once we have a function that specifies the relative weights and the functional interdependencies among constituent ideals, we are able to derive a set of *moral indifference curves*; that is, we are able to rank alternate ways of arranging the basic structure in accordance with the extent to which they achieve the optimal realization of our fundamental ideals.

Precisely specifying Rawls's objective function is too demanding to undertake here; nor is it necessary to make my case. All that's required is to show that Rawls acknowledged that different institutional schemes (constituted by different principles) can realize his fundamental ideals to a greater or lesser extent, that there are some rates at which we should make trades among ideals, or that the set of choice variables exhibit some functional interdependence. If this can be shown, we could, in principle, reverse engineer Rawls's objective function through careful analysis. Showing this might be a little difficult, though, for two reasons. The first is that Rawls orders his principles lexically, repeatedly

<sup>24</sup> Harsanyi (1975).

stating that we are not to tradeoff individuals' rights for more equal opportunity sets or that we are not to tradeoff less equal opportunity sets for greater material prosperity. The second is that Rawls only ever discussed in any detail the relationships that are to hold among his fundamental ideals in the optimal case. Put differently, the functional relationships among ideals modelled by the original position constitute the relationships that are to hold in the case of full justice.<sup>25</sup> Hence, Rawls does not say much about the relative importance of his fundamental ideals; nor does he say much about the rates at which he would be willing to trade off more realization of one ideal for less of another when we are unable to realize both to the optimal degree.

The first difficulty is a red herring. The issue here is whether Rawls acknowledged the possibility of tradeoffs among his *choice variables*—namely, his fundamental ideals, such as equality, respect for persons, rational acceptability, and the like. As I discuss below, his principles of justice and priority rules are meant to specify *optimum conditions*, conditions that must obtain if we are to achieve the optimal realization of his fundamental ideals. There is no inconsistency in acknowledging tradeoffs among choice variables in general while maintaining that the optimal outcome is constituted by tradeoff-forbidding principles. Such a situation would arise if the optimal relationships between choice variables—the relationships that constitute the optimal outcome—constitute a precise balance. Hence, we can acknowledge (along with Rawls<sup>26</sup>) that, under some circumstances, we would be willing to exchange some amount of individual liberty for greater material prosperity while insisting that the optimum conditions—the necessary conditions for realizing a fully just institutional scheme—forbid such tradeoffs.<sup>27</sup>

Although Rawls says little about the relative importance of his fundamental ideals or their interdependence, he does make comments that suggest he is working with an underlying objective function. For example, as I've already noted, he suggests that limiting the basic liberties of some (thereby realizing equality and respect for persons to a lesser degree) for the sake of social and economic gains can be warranted in conditions of extreme scarcity. Further, Rawls suggests that at least some of his fundamental ideals are interdependent; for example, he often claims that the “fair value” of political liberties depends on a distribution of social and economic goods that tends toward equality.<sup>28</sup> Perhaps most telling is the fact that parties in the original position are meant to rank the alternative sets of principles they encounter based on the outcomes these alternatives

25 Cf. Rawls (1999, 104).

26 Rawls (1999, 55, 132).

27 Note further that Rawls's priority rules are derived from—not built into—his model of the optimal balance of fundamental ideals, i.e., the original position. See Rawls (1999, 37).

28 E.g., Rawls (1999, sec. 36).

would engender.<sup>29</sup> The very possibility of ranking alternatives implies the possibility of tradeoffs between choice variables. Different sets of principles are more or less acceptable to individuals situated as they are in the original position. Since the original position models that optimal balance of fundamental ideals, it follows that the outcomes yielded by implementing different sets of principles approximate the optimal balance of ideals to a greater or lesser extent. This implies that the outcomes produced by the principles that encode any particular balance of ideals realizes each fundamental ideal to a greater or lesser extent. Since Rawls supposes that alternatives that realize our fundamental ideals to differing degrees can be ordered, it follows that there is some schedule of rates at which we are willing to make tradeoffs among our ideals. Thus, even though we cannot precisely specify it, Rawls's theoretical exercise seems to presuppose an objective function that assigns relative weights to his constituent ideals and specifies some interdependencies among them.

A constraint function delimits the feasible set of outcomes by specifying the “production possibilities”. This requires specifying the ways in which the value assigned to one variable depends on the value assigned to others. In the present context, such “production interdependencies” arise when the degree to which one ideal is manifest depends on the degree to which other ideals are manifest; for example, the degree of freedom we possess might depend on the extent to which we are physically secure or live within a supportive community. Specifying the feasible set also requires identifying any limits on the values that can be assigned to choice variables; for example, any limits on the extent to which individual liberty or social equality can be realized.

Again, to show that Rawls operated with a constraint function in mind, we need not specify that function precisely; we need only show that he acknowledged limits on the realization of his fundamental ideals or production interdependencies among them. And there's no question that he did. Among others, Rawls acknowledged that the degree to which his fundamental ideals could be realized was limited by moderate resource scarcity, limited altruism, and the fact that individuals hold conflicting conceptions of a worthwhile life.<sup>30</sup> Moreover, Rawls seems to acknowledge various production interdependencies among his ideals when, for example, he claims that the extent to which individuals receive a fair share of social goods consistent with significant scope for individual liberty depends on the extent to which they are treated as equals in collective decision-making

<sup>29</sup> Rawls (1999, sec. 21).

<sup>30</sup> Rawls (1999, sec. 22). Famously, interactions with other societies and shortfalls of compliance with the specified principles are not among the assumed constraints; see (Rawls, 1999, 7, 8). It is the failure of these assumptions to obtain in the actual world that has sparked much criticism of Rawls's ideal theory.

and given effectively equal opportunities.<sup>31</sup> He also suggests that the manifestation of reciprocity and respect for equality depends on the extent to which individuals enjoy effective political liberties.<sup>32</sup> There should be no doubt that Rawls's analytical apparatus includes a constraint function, even if he didn't specify it precisely.

As we saw above, the point of a constrained optimization exercise is to specify a set of conditions that characterize the optimal outcome given the objective and constraint functions. In the case at hand, the task is to characterize the normative political principles that constitute the institutional scheme(s) that realize Rawls's fundamental ideals under the specified constraints. To carry out this task, Rawls constructs the original position to model the balance of ideals that fully just institutions are to manifest and various constraints on the manifestation of those ideals.<sup>33</sup> The original position is simply an inferential device, a means to enable "strictly deductive reasoning" from premises about our ideals and the constraints on their manifestation to the principles of justice that realize a well-ordered society. This deductive reasoning is facilitated by treating principles of justice as the result of an agreement among "rational individuals with certain ends and related to each other in certain ways"; then the principles will (as in rational choice theory) follow deductively from individuals' "beliefs and interests, their situation and the options open to them".<sup>34</sup> The principles derived from the original position deliver

*Rawls's Optimum Conditions.* (1) "Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all." (2) "Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equality of opportunity." (3) The first principle is given lexical priority over the second, so "the basic liberties can be restricted only for the sake of liberty." (4) The fair equality of opportunity principle, (2b), "is prior to the difference principle [2a]". (5) The second principle is lexically prior to "the principle of [Pareto] efficiency and to that of maximizing the sum of advantages".<sup>35</sup>

The two principles and the priority rules (3–5) specify the conditions an institutional

31 Rawls (1999, secs. 12–14, 24).

32 Rawls (1999, 205).

33 Rawls (1999, ch. 3).

34 Rawls (1999, 103).

35 Rawls (1999, 266f).

scheme must satisfy to optimize Rawls's objective function — that is, to achieve the objective of realizing the institutional scheme of a well-ordered society. According to Rawls, these principles represent necessary conditions for a fully just society, specified in terms of the relationships that must hold between constituent ideals (e.g., liberty, equality, welfare).

The point of this section has been to show that the analytical apparatus Rawls uses to specify normative political principles bears all the structural features of a constrained optimization problem. To this end, it is important to note the analogy between Rawls's theoretical exercise and that of welfare economists. Like the theory of Pareto optimality, Rawls's specification of normative principles starts with an objective (a fully just institutional scheme) and a set of choice variables (the fundamental ideals that are more or less realized by alternative institutional schemes). He presupposes an objective function, which specifies the relative contributions made by his fundamental ideals to the moral value of an institutional scheme, as well as the ways in which these contributions are functionally interdependent. Finally, he implicitly relies on a constraint function, which specifies the set of feasible outcomes. The normative principles he derives from the original position thus constitute a set of optimum conditions — that is, a set of conditions that must obtain if we are to achieve our objective of a fully just institutional scheme. If the theory of Pareto optimality is an exemplary case of a constrained optimization exercise, then, by analogy, we can naturally interpret Rawls as solving a constrained optimization problem. Given that Rawls's methodology is representative of a wide range of political theories, we can understand at least some forms of political theory as the sort of optimization exercise to which the theory of second best applies.<sup>36</sup>

#### 4. INTERPRETING THE ANTECEDENT

Let's return to the issue of how to interpret the theorem's antecedent. Recall Goodin's gloss on the theoretical presuppositions version of the theorem: "If any one of the conditions presupposed by ideal theory is missing, then the Theory of Second-Best warns that we might... need to make systematic alterations right across the board in the prescriptions of ideal theory". Let's state this a little more precisely. Let  $C = \{c_1, \dots, c_n\}$  be the set of circumstances assumed by a theorist when specifying normative principles (optimum conditions); let  $P = \{p_1, \dots, p_n\}$  be the set of normative principles specified assuming  $C$ .

<sup>36</sup> To show that the theory of second best applies to normative political theory, I must also show that we get counterintuitive results if we apply ideal normative principles in cases where the theorem's antecedent is satisfied. I leave this until section 6.

*The Theorem of Second Best* — *TP*. If there is some  $c_i \in C$  that can't be satisfied, then, for all  $i$ , we should not necessarily adopt  $p_i \in P$  as a normative guideline for nonideal theory.<sup>37</sup>

This formulation is at least a *prima facie* plausible interpretation of L&L's theorem. Recall L&L's gloss on the antecedent: "if one of the Paretian optimum conditions cannot be fulfilled...". It is perhaps too quick to declare that L&L are referring to what I've called the "Paretian Optimum Conditions" above — the identity statements about marginal rates of substitution and transformation. Their informal statement is ambiguous between that set of derived conditions and the familiar modeling assumptions used to derive those conditions: the absence of production and consumption externalities; that agents have symmetric and perfect information; the absence of monopolies; and costless transactions. If by "Paretian optimum conditions", L&L mean the modeling assumptions ( $C$ ) used to derive the optimum conditions ( $P$ ), then the theoretical presuppositions interpretation is plausible.

A closer look at the theorem's proof shows that L&L's use of "Paretian optimum conditions" does not refer to the modeling assumptions ( $C$ ) but to the three proportionality conditions I've dubbed the "Paretian Optimum Conditions" ( $P$ ). Since this point matters for the theorem's deployment against the ideal guidance view, it's worth carefully walking through the key steps of the proof to see this.<sup>38</sup>

The proof assumes a generic differentiable objective function of  $n$  choice variables,  $F(x_1, \dots, x_n)$ . This function is to be optimized subject to a differentiable constraint on the variables,  $G(x_1, \dots, x_n) = 0$ . Using the Lagrange method, the function  $F$  is at its constrained optimum just in case

$$F_i - \lambda G_i = 0 \text{ for all } i = 1, \dots, n. \quad (1)$$

If we select an arbitrary variable, say  $x_n$ , to serve as the standard of relative value (the "numeraire"), then we can rearrange (1) to yield the  $n - 1$  first-order optimum conditions

<sup>37</sup> Note that, as Lipsey (2007) clearly acknowledges, second best situations need not arise from constraints that are impossible to overcome, but can also arise from policy-created constraints, which in some sense *can* be overcome.

<sup>38</sup> See L&L, sec. 7. There are some typographical errors in the original proof, which are corrected in Lipsey and Lancaster (1997). Ng (2004, ch. 9) presents a more accessible proof, including a helpful graphical illustration of the theorem's key implication.

<sup>39</sup>  $F_i = \frac{\partial F}{\partial x_i}$  is the partial derivative of  $F$  with respect to  $x_i$  and  $G_i = \frac{\partial G}{\partial x_i}$  is the partial derivative of  $G$  with respect to  $x_i$ ;  $\lambda$  is the "Lagrange multiplier".

## THE THEORY OF SECOND BEST

that must hold to attain an efficient allocation:

$$\frac{F_i}{F_n} = \frac{G_i}{G_n} \text{ for all } i = 1, \dots, n-1. \quad (2)$$

Welfare economists interpret  $F$  as a social welfare function and  $G$  as a production constraint. Hence, the optimum conditions depicted in (2) are interpreted as the third Paretian condition above: the marginal rate at which consumers substitute good  $x_i$  for  $x_n$  must be equal to the marginal rate at which the economy can redirect production of  $x_i$  into production of  $x_n$ . But it's important that the theorem's application does not turn on this interpretation. Since  $F$  and  $G$  are both generic functions, these derived conditions more generally represent the optimal relationships among the objective function's constituent variables.

The next step in the proof introduces a second best situation, modeled as a constraint that prevents an arbitrary optimum condition from being satisfied. Since nothing hangs on the choice of condition, we can suppose

$$\frac{F_1}{F_n} = k \frac{G_1}{G_n} \text{ for some } k \neq 1. \quad (3)$$

Again using the Lagrange method, the function to be optimized given this new constraint is

$$F(x_1, \dots, x_n) - \lambda G(x_1, \dots, x_n) - \mu \left( \frac{F_1}{F_n} - k \frac{G_1}{G_n} \right). \quad (4)$$

This is just the original optimization problem with the new constraint appearing as the third term. The proof proceeds to derive the new optimum conditions for an efficient outcome given the introduced constraint—a “second best optimum”<sup>40</sup>—showing that the second best conditions are not identical to the first best conditions given in (2). (We can set the remainder of the proof aside.)

Here, we are interested in the correct interpretation of the theorem's antecedent condition; equation (3) is the key to this. This step in the proof shows that a second best

<sup>40</sup> The phrase “second best” is potentially misleading. Suppose we have an ordinal ranking of all possible outcomes, with  $A$  being the ideal,  $B$  being the next best outcome,  $C$  being the best after  $B$ , etc. The proportionality conditions derived from (4) do not characterize  $B$ , i.e., the next best outcome after the ideal outcome. Instead, they characterize the best outcome attainable under the specified constraints. This is consistent with that outcome being ranked 47th in a complete ordinal ranking. “Second best” simply refers to any situation where the conditions for the first best or ideal optimum cannot be fulfilled. (Thanks to Dave Estlund for bringing this issue to my attention.)

situation is constituted by a barrier to fulfilling one of the derived first-order optimum conditions, whether or not this barrier arises due to a failure to fulfill the circumstances assumed by the initial constraint function. This means that the theorem's consequent applies when there is an obstacle to fulfilling one of the derived optimum conditions, not when one of the circumstances presupposed by the initial optimization fails to obtain. Put in the above terms, the theorem's consequent applies when some  $p_i \in P$  can't be satisfied, not when some  $c_i \in C$  fails to obtain.

We've already seen that some familiar forms of normative political theory have the structure of a general optimization problem (section 3). By parity of reasoning, the consequent of the theorem applies to political theory when there arises a barrier to fulfilling one of the principles of justice derived from the model. Importantly, the consequent does not apply in virtue of a failure to fulfill the ideal circumstances assumed by the specification of ideal principles. For example, Rawls's theory of justice does not satisfy the theorem's antecedent condition simply because the strict compliance assumption fails to obtain. The consequent applies to Rawls's theory only if there is a barrier to fulfilling one of his principles of justice in our world.

Stating the antecedent strictly using the above terms, the theorem says:

*The Theorem of Second Best — Strict.* If there is some  $p_i \in P$  that can't be satisfied, then, for all  $j \neq i$ , we should not necessarily adopt  $p_j \in P$  as a normative guideline for nonideal theory.

The theoretical presuppositions interpretation of the theorem is mistaken.

One might object here that L&L's formalization of the second best problem merely illustrates a range of second best problems. After all, L&L's formulation only attends to "distortion[s] between price and marginal cost in some market(s)" and the "implications this had for pricing rules that ought to be followed by the other, non-distorted, sectors".<sup>41</sup> But, nowadays, economists acknowledge that "distortions can arise for a variety of reasons" and that second best theory applies beyond pricing rules in non-distorted sectors of the economy.<sup>42</sup> Even Lipsey acknowledges many "sources of divergence", that is, many barriers to the realization of a Pareto optimal outcome.<sup>43</sup> Hence, we ought not interpret L&L's theorem literally, as I have here; any optimization problem for which the assumptions of the first best problem can't be satisfied should be regarded as a second

<sup>41</sup> Boadway (1997, 3).

<sup>42</sup> Boadway (1997, 4).

<sup>43</sup> Lipsey (2007).



best problem.<sup>44</sup>

It is true that there are numerous barriers to achieving a Pareto optimal outcome, many of which arise due to the failure of economists' assumptions: the existence of monopolies, incomplete markets, consumption externalities, or imperfect information. Hence, there are as many second best problems as there are barriers to achieving an efficient allocation. But this objection misses the mark. Boadway's and Lipsey's multifarious "sources" of distortion or divergence refer to different obstacles to satisfying one of the conditions for achieving a Pareto optimal outcome, that is, to satisfying some  $p_i \in P$ . Oftentimes, the "source" in question might be a failure of some  $c_i \in C$  to obtain. But it need not be. Moreover, the failure of some  $c_i$  to obtain need not entail a barrier to satisfying some  $p_i \in P$ ; in theory at least, there are ways to achieve a Pareto optimum despite the failure of some circumstance required for a competitive market equilibrium. After all, the first fundamental theorem of welfare economics states that a perfectly competitive market system is only a *sufficient* condition for Pareto efficiency.<sup>45</sup>

The connection between the  $c_i$ s and  $p_i$ s becomes attenuated when we shift to normative political theory, where there are no theorems demonstrating any logical connection between the two. Assuming his theory is correct, there is little *prima facie* reason to think that Rawls's principles fail to apply simply because his model assumes away compliance problems such as tax evasion or a propensity to take advantage of others in a weak bargaining position. More generally, it is not *a priori* true that Rawls's principles cease to apply simply because one of his modelling assumptions fails to obtain. Indeed, there is a presumption that we should fulfill his principles so long as our circumstances permit, whatever they are.

The question of what we should do when our circumstances deviate from those assumed by our constraint function misses the point of the theory of second best. The theory concerns what we should do when we confront a much more general deviation, namely, *when we are prevented from fulfilling one of the necessary conditions for a first-best outcome*, whatever the explanation for this.<sup>46</sup> Put differently, a second best problem arises because some  $p_i \in P$  can't be satisfied, not because some  $c_i \in C$  fails to obtain. In terms of welfare theory, it is specifically concerned with how we should proceed when we are prevented from equalizing the marginal rates of substitution and transformation for a pair of goods, whatever the explanation for that failure. In terms of political theory, it is

<sup>44</sup> This objection comes from an anonymous reviewer.

<sup>45</sup> See (Ng, 2004, ch. 2).

<sup>46</sup> Cf. Lipsey: "A 'second best situation' referred to *any* situation in which the first best was unachievable" (Lipsey, 2007, 352); that is, in which the necessary conditions for realizing the optimum do not obtain.

specifically concerned with how we should proceed when we are prevented from fulfilling an ideal normative principle, whatever the explanation for that failure.

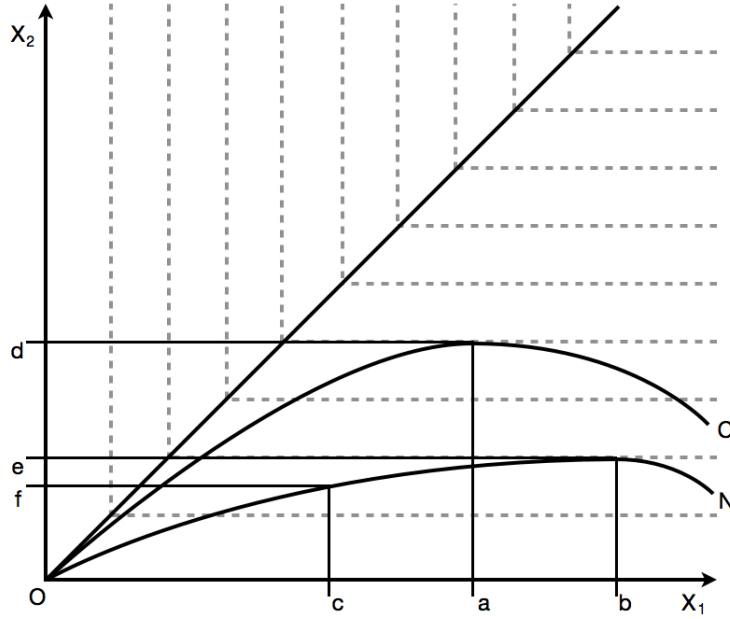
## 5. DISCHARGING THE ANALOGY

I've shown that, interpreted correctly, the theorem of second best says: if an ideal principle can't be satisfied, then the remaining ideal principles do not necessarily characterize an optimal outcome. So far, my argument has been by analogy to the theory of Pareto optimality: (1) The theory of second best proves that, if one of the necessary conditions for a Pareto efficiency can't be satisfied, then the remaining optimum conditions do not necessarily characterize an optimal outcome; (2) Rawls's theory (and, by extension, other political theories with a similar structure) and the theory of Pareto optimality both have the structural features of an optimization problem, with Rawls's normative principles mapping to the necessary conditions for Pareto efficiency; (3) Therefore, the theory of second best proves that, if an ideal principles can't be satisfied, then the remaining ideal principles do not necessarily characterize a just outcome in nonideal circumstances. But one might remain skeptical that the theory of second best has this implication for the ideal guidance view, the structural analogy notwithstanding. Intuitively, one might think that Rawls's difference principle continues to characterize a just institutional structure even if the equal basic liberties principle of the fair equality of opportunity principle can't be satisfied. Why should we think that all of Rawls's principles cease to apply if only one of them cannot be implemented?

In this section, I discharge the analogy, presenting a case that suggests that, if one of Rawls's ideal principles can't be satisfied, then his remaining principles do not necessarily characterize a just institutional structure. To start, consider Figure 1.<sup>47</sup> The  $x$ -axis represents the distributive share of talented and otherwise advantaged individuals, represented by  $X_1$ ; the  $y$ -axis represents the distributive share of individuals with the worst life prospects, represented by  $X_2$ . The 45° line represents all points in the space where the two individuals' shares are strictly equal. Given Rawls's assumption that strict equality is to be preferred unless a departure improves the absolute position of  $X_2$ , the dashed horizontal lines are the moral indifference curves. Let a contribution curve represent the set of feasible distributive shares given some assumptions about the differential rewards that would induce talented and otherwise advantaged individuals to contribute to overall social production. Since  $X_1$  is assumed to be better off, a contribution curve lies below the 45° line everywhere except the origin.

47 Cf. Rawls (1999, 66).

## THE THEORY OF SECOND BEST



**Figure 1.** The difference principle

Let the curve  $OC$  be the *compliance contribution curve*, the curve that would obtain were Rawls's equal basic liberties principle to obtain. The equal basic liberties principle requires that each person enjoy the most extensive package of basic rights consistent with everyone enjoying the same package of rights (Rawls, 1999, 266). The difference principle requires an institutional scheme that yields the distributive shares represented by the point where  $OC$  is tangent to the highest indifference curve; in this case, the point where  $X_1$  receives  $a$  and  $X_2$  receives  $d$ . This distribution is morally preferred to strict equality—which is realized at  $O$ —because it yields a greater absolute share for  $X_2$ . Any point beyond  $a$  on the  $x$ -axis is unjust.

The difference principle specifies the ratio  $a/d$  as the optimal limit on permissible inequality when the equal basic liberties principle is satisfied. But suppose there are barriers to satisfying the equal basic liberties principle. Failure to satisfy the equal basic liberties principle will likely result in a different contribution curve. This *noncompliance contribution curve* might take the shape of  $ON$  (see Figure 1), for several reasons. If

some are granted the right to vote and others are excluded, the former will be able to shape policy to their (economic) advantage; if the rich are granted freedom of speech and assembly and the poor are not, the rich will have a great advantage in lobbying politicians to favor their interests; if some are given freedom from oppression and assault and other are not, the former will be permitted to bully the latter into serving their interests. These rights disparities have two potential consequences. First, they likely diminish the amount of social wealth genuinely available to the worst-off; this reduces their expected prospects, as indicated by the difference between  $d$  and  $e$ . Second, and relatedly, rights disparities allow the better-off to demand a greater share of the total social product in exchange for their contributions to social production. This is represented by the difference between  $b$  and  $a$ . At the point at which the noncompliance curve is tangent to the highest indifference curve,  $X_1$  receives  $b$  and  $X_2$  receives  $e$ . The difference principle permits departures from strict equality up to the point  $b/e$  when the equal basic liberties principles is not fulfilled.

From Figure 1, we see that the extent to which implementing the difference principle fulfills our commitment to values like impartiality, equality, and respect for persons depends on whether individuals enjoy roughly equal packages of rights. Whether the difference principle realizes our fundamental ideals to the fullest extent when the equal basic liberties principles remains unsatisfied depends on how the distributive consequences of implementing that principle compare to the consequences of implementing some other distributive principle. It seems *prima facie* implausible that, while the equal basic liberties principle remains unsatisfied, the distributive principle that best realizes our fundamental values would permit increased inequality, as the difference principle does. This intuition is strengthened once we account for the corrosive political effects of great inequality. Plausibly, the best principle for regulating inequalities when individuals enjoy such disparate packages of rights would locate the limit on inequality somewhere other than  $b/e$ —perhaps at the point along the  $x$ -axis represented by  $c$ .<sup>48</sup> Such a principle decreases  $X_2$ 's prospects slightly in exchange for a drastic reduction in inequality. A principle that locates the limit here can be given a reasonable justification: given that individuals do not enjoy similar packages of rights, we must limit permissible inequalities to  $c/f$  to prevent the better-off from acquiring the additional advantages that come with great relative wealth.

To be clear, the foregoing does not show that, necessarily, we should reject the difference principle if the equal basic liberties principles can't be satisfied. Rather, I've shown

<sup>48</sup> The content of such a principle doesn't matter here, so long as there is a sensible way to give it content. What matters is that the principle locates the limit on permissible inequality somewhere other than  $b/e$ .

that, if the equal basic liberties principle can't be satisfied, then we should not necessarily implement the difference principle. Whether we should depends on the extent to which our fundamental ideals are realized by the outcomes engendered by implementing the difference principle in the absence of equal basic liberties and whether the outcomes engendered by implementing a different distributive principle better realize our ideals.

If there are barriers to satisfying an ideal principle, why, in general, should we be dubious that the remaining ideal principles continue to constitute an optimal outcome? The reasoning underwriting the theory of second best gives us the answer. When we fail to fulfill the necessary condition for an ideal outcome, our attempts to attain a second best optimum must account for complex relationships among the variables in our objective function. In welfare economics, second best conditions must account for consumption and production *interdependencies*; they must account for the complex ways in which individuals' utility from consuming certain goods depends on their consumption of other goods, or the complex ways in which the production of certain goods depends on the production of other goods.<sup>49</sup> When these relationships do not satisfy a stringent separability condition,<sup>50</sup> then we cannot assume that second best conditions will deviate from first best conditions in any straightforward fashion.<sup>51</sup>

In terms of political theory, our attempts to attain a nonideal optimum — an outcome that realizes our fundamental ideals to the fullest extent possible under actual circumstances — must account for interdependencies among the ideals that constitute the moral value of a political society, that is, the ideals that enter our objective function. I mentioned two types of interdependence in section 3, both of which are relevant here.<sup>52</sup> When our ideals are interdependent in either of these ways, and these interdependencies are not separable (e.g., our willingness to exchange some amount of liberty for equality does not depend on the extent to which we realize physical security), then we can't say anything in general about how nonideal political principles compare with ideal political principles. Complex interdependencies prevent us from determining a priori whether ideal principles provide normative guidelines for our efforts to address injustice when there are barriers to satisfying one of those ideal principles. More importantly, there are

49 Cf. Ng: "[T]he second-best conditions depend not only on the values of (ratios of) marginal costs and marginal rates of substitutions, but also on the degrees of complementarity or substitutability between goods in the constrained sector and those in the free sector, and the effects of increased production of a good on the marginal costs of another" (Ng, 2004, 192).

50 See Blackorby, Davidson and Schworm (1991).

51 Roughly, a pair of variables,  $x, y$ , is separable in the sense relevant here if the marginal rates of substitution between  $x$  and  $y$  do not depend on a third variable,  $z$ . Satisfying this condition is quite unrealistic in any normal economic setting.

52 Goodin (1995, pp. 54–55) presents additional examples.

good reasons to think that ideal principles cease to constitute an optimal state of affairs when there is a barrier to satisfying one of them. Hence, if an ideal principle can't be satisfied, we should not necessarily seek to satisfy the remaining ideal principles, even if we can do so.

## 6. WHY THE ANTECEDENT MATTERS

The discussion in the section 4 is not mere pedantry. How we interpret the theorem's antecedent is consequential for deploying it against the ideal guidance view. Consider the standard argument against the ideal guidance view delivered by the theoretical presuppositions interpretation of the theorem:

- (1) *Definition of Ideal Theory.* Let  $C = \{c_1, \dots, c_n\}$  be the set of circumstances an ideal theorist assumes when specifying  $P = \{p_1, \dots, p_n\}$ , the set of ideal principles. There is at least one  $c_i \in C$  that can't be satisfied in the actual world.
- (2) *The Theorem of Second Best—TP.* If there is some  $c_i \in C$  that can't be satisfied, then, for all  $i$ , we should not necessarily adopt  $p_i \in P$  as a normative guideline for nonideal theory.
- (3) Therefore, for all  $i$ , we should not necessarily adopt  $p_i \in P$  as a normative guideline for nonideal theory.

This argument apparently delivers a powerful challenge to the ideal guidance view because it need only appeal to a relatively uncontroversial definition of ideal theory and an allegedly deductively proven theorem. But, as I show in section 4, L&L do not prove premise (2). Moreover, this argument ceases to be valid if we deploy the strict version of the theorem, substituting (2) for (2').

- (2') *The Theorem of Second Best—Strict.* If there is some  $p_i \in P$  that can't be satisfied, then, for all  $j \neq i$ , we should not necessarily adopt  $p_j \in P$  as a normative guideline for nonideal theory.

So the standard argument cannot rest on the rigor of L&L's proof. The standard argument might nonetheless be sound. But showing this requires an argument for premise (2), an argument philosophers have declined to provide (since they have deferred to L&L at this point). Moreover, intuition and conventional wisdom seem to oppose premise (2); many philosophers have argued that ideal principles continue to provide normative guidelines for addressing injustice despite the failure of ideal circumstances to obtain in the actual

world.<sup>53</sup> The challenge to the ideal guidance view posed by the standard argument has yet to be vindicated.<sup>54</sup>

A valid argument that deploys the theorem of second best must take a stance on the feasibility of implementing ideal principles in the actual world:

- (1') There is at least one  $p_i \in P$  that cannot be satisfied in the actual world.
- (2') *The Theorem of Second Best—Strict*. If there is some  $p_i \in P$  that can't be satisfied, then, for all  $j \neq i$ , we should not necessarily adopt  $p_j \in P$  as a normative guideline for nonideal theory.
- (3') Therefore, for all  $j \neq i$ , we should not necessarily adopt  $p_j \in P$  as a normative guideline for nonideal theory.

Premise (1') is controversial, to say the least. Whether ideal principles can be implemented in the actual world and, hence, can serve as useful guides for nonideal circumstances is precisely what's at issue between ideal guidance skeptics and proponents. As of yet, there is no compelling argument to show that implementing, say, Rawls's principles of justice is infeasible in our circumstances.<sup>55</sup> Mere suggestion that an ideal principle might not be feasible to implement is likely not enough to trigger the theorem of second best. At least, the mere suggestion of infeasibility will be insufficient to persuade many philosophers, most of whom think that the burden of proof surely rests with the feasibility skeptic.

In sum, the ideal guidance skeptic has some work to do. On the one hand, ideal guidance skeptics typically focus on the question of whether principles specified under idealized assumptions can provide guidelines for nonideal theory, appealing to the theory of second best to bolster their denial that they can. But, contrary to popular belief, L&L's theorem does not address this issue; L&L do not prove premise (2) in the standard argument. The theory of second best speaks to a different issue, namely, whether we should aim to approximate what remains of an ideal theory if one of the ideal principles remains unsatisfied. Hence, if the standard argument is to pose a challenge for the ideal guidance view, premise (2) requires an argument yet to be provided. On the other hand, the argument deploying the strict version of the theorem does potentially pose a powerful challenge to the ideal guidance view insofar as ideal theory generally yields principles that cannot be implemented in the actual world. But whether ideal theory generally

<sup>53</sup> See, e.g., Lawford-Smith (2010); Mason (2004); Simmons (2010); Valentini (2009).

<sup>54</sup> Although Wiens (forthcoming) goes some distance in this direction.

<sup>55</sup> Though there are plenty of arguments meant to show that implementing Rawls's principles in our circumstances is *undesirable*; see, e.g., Farrelly (2007); Mills (2005).

yields principles that can't be implemented is highly controversial; as yet, we lack a compelling argument to support premise (1'). Worse (for the ideal guidance skeptic), it seems nigh impossible to persuasively demonstrate premise (1') short of actually failing to successfully implement a sufficiently wide range of ideal principles despite trying to do so. Thus, the theory of second best poses little challenge to the ideal guidance view, even if it challenges the guidance capacity of a particular ideal theory. Ideal guidance skeptics would do best to abandon appeal to the theory of second best in their efforts to deny the ideal guidance view.

#### REFERENCES

- Blackorby, Charles, Russell Davidson and William Schworm. 1991. "The Validity of Piece-meal Second-Best Policy." *Journal of Public Economics* 46:267–290.
- Boadway, Robin. 1997. The Role of Second-Best Theory in Public Economics. In *Trade, Technology and Economics: Essays in Honour of Richard G. Lipsey*, ed. B.C. Eaton and R.G. Harris. Cheltenham, UK: Edward Elger pp. 3–25.
- 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.
- Farrelly, Colin. 2007. "Justice in Ideal Theory: A Refutation." *Political Studies* 55(4):844–864.
- 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.
- Goodin, Robert E. 2012. The Bioethics of Second Best. In *Global Justice and Bioethics*, ed. Joseph Millum and Ezekiel J. Emanuel. New York: Oxford.
- Hamlin, Alan and Zofia Stemplowska. 2012. "Theory, Ideal Theory and the Theory of Ideals." *Political Studies Review* 10:48–62.
- Harsanyi, John. 1975. "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory." *The American Political Science Review* 69(2):594–606. John Rawls by A Theory of Justice.



## THE THEORY OF SECOND BEST

- Lawford-Smith, Holly. 2010. "Ideal Theory — A Reply to Valentini." *Journal of Political Philosophy* 18(3):357–368.
- 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.
- Mason, Andrew. 2004. "Just Constraints." *British Journal of Political Science* 34:251–268.
- Mills, Charles W. 2005. "'Ideal Theory' as Ideology." *Hypatia* 20(3):165–184.
- Ng, Yew-Kwang. 2004. *Welfare Economics: Towards a More Complete Analysis*. New York: Palgrave Macmillan.
- Rawls, John. 1999. *A Theory of Justice*. 2 ed. Cambridge, MA: Harvard University Press.
- 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.
- Swift, Adam. 2008. "The Value of Philosophy in Nonideal Circumstances." *Social Theory and Practice* 34(3):363–387.
- Valentini, Laura. 2009. "On the Apparent Paradox of Ideal Theory." *Journal of Political Philosophy* 17(3):332–355.
- Valentini, Laura. 2011. "A Paradigm Shift in Theorizing About Justice? A Critique of Sen." *Economics and Philosophy* 27:297–315.
- Valentini, Laura. 2012. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7(9):654–664.
- Wiens, David. 2012. "Prescribing Institutions Without Ideal Theory." *The Journal of Political Philosophy* 20(1):45–70.
- Wiens, David. forthcoming. Achieving Global Justice: Why Failures Matter More Than Ideals. In *Making Global Institutions Work: Power, Accountability and Change*, ed. Kate Brennan. New York: Routledge.