Value Invariabilism and Two Distinctions in Value

Abstract (298 words): Following Moore, value invariabilists (conditionalists) deny that something’s intrinsic value can be affected by features extrinsic to it. In this paper, I do four things. First, I reconceptualize the disagreement between variabilists and invariabilists in light of a contemporary distinction between intrinsic value and final value. Making this distinction helps clarify why invariabilism is trivially true in one sense (intrinsic value) while maintaining that there is a robust debate in another sense (final value): while something’s intrinsic value—i.e., the value that depends solely on its intrinsic nature—cannot vary in different contexts on pain of contradiction, there’s room for debate about whether the value something contributes to the whole in which it’s a part (i.e., its final value) can vary. Second, I apply my reconceptualization of the variabilism debate to one of the primary arguments offered for invariabilism about final value simpliciter (offered by Lemos) and argue that, while it fails, the argument supports invariabilism about final welfare value. The rest of the paper details challenges to this argument. First, (i.e., section III) I address a direct challenge to the claim that final welfare value invariabilism is a better explanation than final value simpliciter invariabilism. I consider two objections that have been raised to the explanation offered in section II—one of theoretical parsimony and one involving desire-satisfaction accounts of welfare—and argue that we should reject both. Next (i.e., section IV), I consider how one might leverage invariabilism about final welfare value to establish invariabilism about final value simpliciter: a direct and proportionate correspondence of welfare value to value simpliciter in conjunction with my endorsement of final welfare invariabilism entails final value simpliciter invariabilism. In response, I argue that this correspondence is less plausible than an incompatible principle about trade-offs in welfare and final value simpliciter.

Word count: 4837

Keywords: Value Theory, Metaethics, Invariabilism, Welfare, Final Value
I. Introduction

Fletcher (2010) identifies value atomism with two claims:

- Invariabilism: for all $\Phi$, the intrinsic value of $\Phi$ is invariant to changes in context (i.e., features extrinsic to $\Phi$ cannot affect $\Phi$’s intrinsic value).
- Additivity: the intrinsic value of a whole composed of $\Phi$ and $\Psi$ (where $\Phi$ and $\Psi$ don’t overlap) is equal to the sum of $\Phi$’s and $\Psi$’s respective intrinsic values.

Value holism (i.e., the denial of value atomism) requires denying at least one of these claims (Fletcher 2010: 162). Thus, there are three holist possibilities:

(i) invariabilism is true but additivity is false;
(ii) additivity is true but invariabilism is false;
(iii) both additivity and invariabilism are false.

Moore was a (i) holist. Some contemporary value theorists follow Moore in endorsing invariabilism and rejecting additivity. Others have rejected invariabilism. In this paper, I examine one of the primary arguments offered in support of invariabilism. First (section II) I precisify the formulation of invariabilism and the debate with contemporary meta-axiological distinctions. In section III, I explain one of the primary arguments in favor of invariabilism and argue that, while it fails to establish its target, it does support a different conclusion about invariabilism. Section IV addresses some anticipated objections Lemos (1998) raises. Finally, I consider an argument leveraging my conclusion in section III to establish Lemos’s target conclusion and find it wanting.

---

1 See also Oddie (2001) and Brown (2007).
II. Invariabilism Reconceptualized

In this section, I reconceptualize the disagreement about invariabilism. Invariabilism, as formulated above, seems analytically true: if we think of intrinsic value as value that solely depends on intrinsic properties, denying invariabilism appears incoherent, for we should expect that same value whenever we see those intrinsic properties instantiated. My reconceptualization respects this intuition and reframes the debate in terms of final value. In what follows, I highlight some distinctions made in contemporary value theory and show how they help illuminate the invariabilism/variabilism debate.

The first set of distinctions I’ll discuss originate in Korsgaard (1983):

*Final Value*: \( \Phi \) is finally valuable if and only if it is valuable as an end.

*Instrumental Value*: \( \Phi \) is instrumentally valuable if and only if it is valuable as a means.

*Intrinsic Value*: \( \Phi \) is intrinsically valuable if and only if it is valuable in a way that supervenes only on the intrinsic properties of \( \Phi \).

*Extrinsic Value*: \( \Phi \) is extrinsically valuable if and only if it is valuable in a way that supervenes – at least to some extent – on the extrinsic properties of \( \Phi \). (Dorsey 2012a: 137).

For philosophers who endorse the distinction, a primary motivation for drawing a distinction between final value and intrinsic value is that, for at least some \( \Phi \), it appears \( \Phi \)’s final value supervenes on properties extrinsic to \( \Phi \). Examples include the pen Abraham Lincoln used to sign the Emancipation Proclamation, Princess Diana’s dress, and a mink coat. Though these examples are so common in the literature that “many now merely mention them to establish the distinction between final and intrinsic value,” Tucker argues that they fail to motivate the distinction (2016:

---

5 E.g., Korsgaard (1983); Kagan (1998); Rabinowicz and Rønnow-Rasmussen (2000); Olson (2004); Dorsey (2012b); Hurka (1998); Fletcher (2009); Bradford (2013).


Other philosophers think that the final value of pleasures and pains provide “the clearest way of articulating the [variabilism/invariabilism] debate” (Olson 2004: 37).9

Before I introduce an amendment to the above conception of final value, it’s useful to make a distinction between well-being and value simpliciter. Well-being concerns which things are good or bad for us, what makes one better or worse off, what makes one’s life go well or worse for them, etc.10 Value simpliciter is different; when making judgments about the value simpliciter of a state of affairs, “we are not saying that the state of affairs is good or bad for someone… we are saying that it is simply a good or a bad situation” (Heathwood 2015: 137). Some judgments, such as “it’s a bad thing that the wicked are well off,” involve both axiological concepts (Heathwood 2015: 137). First, the claim is about something that is bad simpliciter: that the wicked are well off is simply a bad state of affairs. But, the state of which the bad-simpliciter claim is about concerns well-being: the wicked are well off, i.e. the state of affairs is good for the wicked. So, while the state of affairs might have positive well-being value, it might be disvaluable simpliciter. (In Section IV and V, I discuss objections to my argument in Section III that attempt to reduce well-being to value simpliciter and vice versa, respectively; my current invocation of the distinction is aimed at illuminating the concept of final value).

With the well-being/value simpliciter distinction in hand, I propose we amend the concept of final value. This amendment is motivated by an intuitive answer to a question: what is it for something to be valuable as an end for a person (i.e. finally welfare valuable)? My intuitive answer is to identify this with a certain role: directly making a person better off. As something that is instrumentally welfare valuable is something that indirectly makes a person better off—in virtue of bringing about something else that’s finally welfare valuable—it’s natural to think that something is

---


10 In what follows, I use “well-being”, “welfare”, and “prudential value” interchangeably.
finally welfare valuable when it *directly* makes a person better off. So, when a state of affairs is valuable as an end *for a person*, its obtaining directly increases that person’s level of well-being.

In light of this, I propose that $\Phi$ is finally welfare valuable for $S$ iff $\Phi$ directly makes $S$ better off (i.e. directly contributes to a higher welfare value for $S$). Now, well-being concerns states of affairs that are good for a person while value *simpliciter* concerns those that are simply bad. Rather than a person being the relevant metric, then, the relevant metric for value *simpliciter* is at the state of affairs of which it’s a part. So, applying my amendment to final welfare value to value *simpliciter* gives us:

For a state of affairs $\Phi$, which is part of a larger state of affairs $S$, $\Phi$ is finally valuable *simpliciter* in $S$ to degree $D$ iff $\Phi$ directly makes $S$ more valuable (i.e., directly improves the value of $S$) by $D$.$^{11}$

Now, the distinction between final and intrinsic value may complicate how we ought to understand invarielabilism. If final value and intrinsic value are distinct and not necessarily coextensive, there are four variabilism/invarielabilism combinations:

(1) **Strong Invarielabilism**: For all $\Phi$, features extrinsic to $\Phi$ cannot affect $\Phi$’s intrinsic value or $\Phi$’s final value

(2) **Final Value Variabilism**: For some $\Phi$, features extrinsic to $\Phi$ cannot affect the intrinsic value of $\Phi$ but can affect the final value of $\Phi$.

(3) **Intrinsic Value Variabilism**: For some $\Phi$, features extrinsic to $\Phi$ cannot affect the final value of $\Phi$ but can affect the intrinsic value of $\Phi$.

(4) **Strong Variabilism**: For some $\Phi$, features extrinsic to $\Phi$ can affect both $\Phi$’s intrinsic value and $\Phi$’s final value.$^{12}$

$^{11}$ My conception of final value is similar to the one Bradley (2002) discusses. For other conceptions of final value and criticism of them, see Tucker (2016).
Intrinsic Value Variabilism and Strong Variabilism are problematic for the same reason: assuming the prior conception of intrinsic value (i.e., that $\Phi$’s intrinsic value is value that supervenes only on the intrinsic properties of $\Phi$), these are incoherent. Suppose that $\Phi$’s intrinsic value is value that supervenes solely on $\Phi$’s intrinsic properties. If so, then it’s impossible for features extrinsic to $\Phi$ to affect $\Phi$’s intrinsic value. But, according to Intrinsic Value Variabilism and Strong Variabilism, it’s possible for features extrinsic to $\Phi$ to affect $\Phi$’s intrinsic value. Thus, one must either abandon these theses or this conception of intrinsic value.

Some abandon this thin conception of intrinsic value.\textsuperscript{13} I find this infelicitous: if $\Phi$’s intrinsic value is not value that depends solely on $\Phi$’s intrinsic properties, I lose track of what work ‘intrinsic’ supposedly does. Perhaps stipulation is needed to avoid merely semantic disputes. Here’s my modest, intuitive solution: let’s accept the thin conception of intrinsic value—value that supervenes only on something’s intrinsic properties—and that final value is value as an end (i.e., value that directly improves the value for a person or the world). Rather than debating the distinction’s legitimacy, accepting it allows us to refocus the debate on whether or not final value is necessarily coextensive with intrinsic value. If so, then we need only discover whether something is intrinsically valuable and to what degree to determine its final value. If final value isn’t necessarily coextensive with intrinsic value, then we must account for which features or other states of affairs affect something’s final value and how much. The standard cases, such as Diana’s dress and Lincoln’s pen, are properly interpreted as counterexamples to the claim that final value and intrinsic value are necessarily coextensive. While not ground-breaking, this suggestion might help with the linguistic

\textsuperscript{12} An astute eye will notice I’ve dropped the well-being/value \textit{simpliciter} distinction here; it thickens the conceptual scheme. I address this in Section III.

\textsuperscript{13} Tucker notes Bradley (2002), Dorsey (2012b), Hurka (1998) and Kagan (1998) as those who use “intrinsic value” to “refer to the value something has as an end, or for its own sake” and use something else for value in virtue of intrinsic features (2016: 1912 fn1).
Wild West we find ourselves in. And so, adopting this proposal, we must reject the Intrinsic Value Variabilism and Strong Variabilism.

I’ve argued that final value is value that directly makes a person or world better off, which helps clarify why I take the primary disagreement between the invariabilist and the variabilist to be between Strong Invariabilism and Final Value Variabilism. Both must accept, given the thin conception of intrinsic value, that only Φ’s intrinsic features can affect Φ’s intrinsic value. By affirming Strong Invariabilism, the invariabilist asserts that the amount of value that Φ directly adds to a person’s life or the world depends only on features intrinsic to Φ, and so the amount of value Φ directly contributes is the same in all contexts. The variabilist, on the other hand, thinks that the amount of value that something directly contributes to the world or a person’s life can be affected by extrinsic features. This is captured by Final Value Variabilism. Thus, the variabilism/invariabilism debate is a disagreement between whether Strong Invariabilism or Final Value Variabilism is true.

III. Why A Primary Argument for Strong Invariabilism Fails

In this section, I first explicate one of the primary arguments for invariabilism—offered by Lemos—in light of section II. Then, I argue that it fails to establish its target (Strong Invariabilism for final value *simpliciter*) but succeeds in establishing Strong Invariabilism for final well-being.

First, consider the following states of affairs:

(1) $A$ is pleased and $A$ is wicked.

(2) $A$ is pleased.

Strong Invariabilists about final value *simpliciter* will hold that $A$’s being pleased has the same final value *simpliciter* in (2) as it does in (1); Final Value Variabilists about final value *simpliciter* will hold that $A$’s being pleased in (1) is at least less finally valuable than it is in (2), because they think $A$’s being
wicked affects the final value simpliciter of A’s being pleased in (1).  

(Some philosophers have criticized the variabilist’s interpretation of this case. I will put these aside to focus on one of the primary positive arguments offered for invariabilism being true). The overall structure of the argument intended to directly support invariabilism is offered by Lemos, and can be formalized as follows:

(P1) A state of affairs in which an individual is both vicious and pleased is a fitting object of indignation.

(P2) The best explanation for why (P1) is true is invariabilism about final value simpliciter.

(P3) We should believe the best explanation.

(C2) Therefore, we should believe that invariabilism about final value simpliciter is true.

Formalized this way, Lemos’s discussion motivates the truth of (P2). Following the line of reasoning above, variabilists will argue that (P2) is false (because, e.g., the best explanation for (P1) being true is that the individual being wicked makes their pleasure finally bad simpliciter). To defend (P2), Lemos argues that what makes cases like those in (P1)—e.g., our case (1)—fitting objects of indignation is that the wicked person “has a good [being pleased] that he deserves not to have” (1994: 44, italics in original).

First, Lemos compares (1) to another state of affairs:

(3) A is wicked and A is blond.

(3) is not fittingly offensive like (1), and Lemos claims that this is best explained by the fact that A’s being pleased is finally valuable simpliciter while A’s being blond is neither finally good nor bad.

---

14 See, e.g., Olson (2004).
15 See, e.g., Bradley (2002) and Tucker (2016; forthcoming)
16 See, e.g., Olson (2004).
17 The following paragraph is a paraphrase of Lemos (1994: 44-45).
Some might think, Lemos considers, that the wicked having something they don’t deserve to have explains the offensiveness of (1). Citing Aristotle, Lemos distinguishes between unmerited good fortune and unmerited bad fortune. Both unmerited good fortune and unmerited bad fortune are cases in which someone has something they don’t deserve to have, but these are fitting objects of different attitudes: unmerited good fortune is a fitting object of indignation; unmerited bad fortune is a fitting object of pity. Lemos claims that what best explains why the person having unmerited good fortune is a fitting object of indignation is not merely that the she has something she doesn’t deserve to have, but rather that she has something finally good \( \text{simpliciter} \) that she doesn’t deserve to have.\(^{19}\) Thus, Lemos concludes, what makes (1) finally bad \( \text{simpliciter} \) is precisely that \( A \) is wicked and that \( A \) has something finally good \( \text{simpliciter} \) (i.e., being pleased). And, the reason for this is that \( A \)’s being pleased retains its goodness, which supports invariabilism being true.

(A quick clarification: though Lemos frames his argument in terms of intrinsic value, a survey of his other work reveals that Lemos doesn’t attend to the intrinsic/final distinction. For example, Lemos understands what I’ve called variabilism as the claim that “a state of affairs that has a certain intrinsic value can have a different intrinsic value when certain other states of affairs obtain” (2015: 135). Given the thin conception of intrinsic value that I’m assuming, Lemos here should be understood as discussing final value \( \text{simpliciter} \). In light of this and my explication of the disagreement between the invariabilist and the variabilist as being between Strong Invariabilism and Final Value Variabilism, we should understand Lemos’s argument as being in favor of the invariabilism of final value \( \text{simpliciter} \), i.e. Strong Invariabilism. Thus, we should understand him as arguing that the final value of (2) doesn’t change when it’s a part of (1)).

\(^{18}\) An astute reader will note the introduction of \( \text{simpliciter} \) here. Lemos has confirmed that his argument is about value \( \text{simpliciter} \) and not well-being in correspondence.

\(^{19}\) Though he doesn’t explicitly do so, Lemos could say something isomorphic regarding unmerited bad fortune.
I can now level my criticism: Lemos’s argument attempts to establish Strong Invariabilism for final value *simpliciter*, but his argument is best understood as supporting Strong Invariabilism for final welfare value. First, consider the blondness argument. Lemos argues that (3) does not fittingly invoke the same indignation as (1) because, while $A$’s being pleased in (1) is finally valuable *simpliciter*, $A$’s being blond in (3) is not. An obvious explanation for this is that $A$’s being pleased in (1) is finally well-being enhancing; $A$’s being pleased in (1) directly makes $A$’s life go better for $A$, and this is fittingly offensive because $A$ is wicked. This explains the difference in our proper evaluations of (1) and (3) as Lemos seems to be suggesting: there is something unfitting when the wicked have lives go well for them because they have something that directly increases their well-being, but being blond is not something that directly increases their well-being and so is not fittingly offensive. This also applies to his argument invoking Aristotle on unmerited good fortune. Here Lemos argues that the explanation for why the person having unmerited good fortune is a fitting object of indignation is not that the she has something she doesn’t deserve to have, but that she has something finally good *simpliciter* that she doesn’t deserve to have. Once again, it seems the proper object of the indignation is that the wicked person has something that directly makes her better off, rather than that she has something that directly improves the value of the world. If so, then this supports Strong Invariabilism about final welfare value and not of final value *simpliciter*. (Indeed, one might think that having good fortune might seem to simply mean that their life is going well for them).

Understanding Lemos’s argument as supporting Strong Invariabilism about final well-being, then, seems to provide a natural explanation of the case, but is it a better explanation than one in which $A$ has something that is finally valuable *simpliciter*? To check, we need a case in which the

20 Of course, if Rod Stewart’s song “Blondes (Have More Fun)” is accurate, being blond might be instrumentally good for a person.
wicked person has something that is finally valuable *simpliciter* but which does not directly make her finally better off. We can construct such a case by making some assumptions.

First, let’s assume a resonance constraint on well-being. This is to say that something can be directly good for a person only if that person has some relevant pro-attitude towards it. Railton famously writes:

> What is [finally] valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone’s good to imagine that it might fail in any such way to engage him. (1986: 9).

The resonance constraint is not uncontroversial. It’s useful to assume here, though, because it provides a way of separating something with final value *simpliciter* from something with final well-being value: if the resonance constraint is not met, then it’s possible that a person has something with final value *simpliciter* that does not directly make her life go better for her.

Suppose, in addition, that understanding has final value *simpliciter*. Unlike the resonance constraint assumption, it’s arbitrary that I’ve chosen an understanding-based account of epistemic value; I invite the reader to substitute their preferred account of epistemic value (other than a nihilistic account). Now, consider the following state of affairs:

(4)  *B* is wicked, understands the crucial importance of bees in our ecosystem, and doesn’t care about understanding this at all.

By the resonance constraint assumption, *B*’s understanding has no final well-being value for *B* because *B* doesn’t care about understanding why bees are crucially important in our ecosystem. By

---

21 E.g., Rosati (1996) and Dorsey (2017) argue in favor of the resonance constraint; Dorsey (2017) and Sarch (2011) criticize Rosati’s arguments; Sarch (2011) also argues against Rosati’s formulation of the resonance constraint.
the assumption that understanding has final value *simpliciter*, (4) has final value *simpliciter*. And so, we have a state of affairs with final value *simpliciter* and without final well-being.

If Lemos’s argument for why (3) is not offensive while (1) is should be understood as A’s being pleased having final value *simpliciter* rather than having something that is finally good for A (i.e., rather than final well-being), then we should find (4) offensive like (1) and unlike (3). After all, if Lemos’s argument is about invariabilism for final value *simpliciter*, then states of affairs in which the subject is wicked and has something finally valuable *simpliciter* should evoke fittingly negative attitudes; in both (1) and (4), the subjects have something that is finally valuable *simpliciter*. But (4) is not offensive in the same way that (1) is. And so it doesn’t seem that the wicked having something with value *simpliciter* captures what makes our negative attitudes towards (1), but rather that the wicked has something that directly makes them better off, i.e. has final welfare value. Thus, Lemos’s argument here appears to be better explained as an argument supporting Strong Invariabilism for final well-being than supporting Strong Invariabilism for final value *simpliciter*.

Thus, I conclude that Lemos provides a strong case for the truth of Strong Invariabilism is true for final well-being but not for the truth of Strong Invariabilism for final value *simpliciter*. I find these arguments convincing but, as I’ve argued, they do not properly defend Strong Invariabilism for final value *simpliciter*.

### IV. Lemos’s Anticipated Responses

I’ve argued that Lemos’s arguments support Strong Invariabilism for final well-being rather than Strong Invariabilism for final value *simpliciter*, because A has something finally prudentially good for A that is contrary to A’s desert in (1) rather than A having something that is finally value *simpliciter* contrary to A’s desert. Lemos (1998) anticipates that one might be drawn to the explanation that I’ve offered and provides two objections to it. Lemos’s anticipated objections, then, would threaten
my argument in section III. I think that there are good reasons to doubt both his points and will discuss these in this section.

First, Lemos argues that this explanation fails to explain why indignation is the fitting response rather than “pity or curiousity” on at least some conceptions of what it is for something to be welfare valuable (1998: 335). Lemos considers a desire satisfaction account of well-being (i.e., $X$ is good for $S$ just in case $S$ desires $X$ for its own sake) and an idealized version (i.e., $X$ is good for $S$ just in case $S$ would desire $X$ were $S$ ‘fully informed’) and describes a type of case in which (i) a person is treated contrary to his desert and (ii) that treatment has welfare value for him because he desires it for its own sake (or would desire it were he ‘fully informed’):

Now imagine that it is contrary to $S$’s desert that he be treated in a certain way, e.g. that he be punished for some crime he didn’t commit or beaten by thugs. But suppose that $S$ happens to desire that sort of thing for its own sake or would desire it were he ‘fully informed.’ In such a case, $S$’s being punished would seem to be (a) contrary to his desert and (b) ‘good for him’” (Lemos 1998: 335).

By the theories of well-being Lemos considers here, these cases are like (1) in that they are states of affairs that are final prudentially valuable for their subjects ($A$ and $S$, respectively) and that involve their subjects having something contrary to what they deserve ($A$ being pleased and $S$ being undeservedly punished or beaten, respectively). But, while (1) seems to be a fitting object of righteous indignation, these cases don’t; rather, these cases seem to be fitting objects of “our pity or curiousity” (1998: 335). And so, it can’t be that the best explanation for our righteous indignation towards (1) is best explained by someone having something good for them that they don’t deserve to have, for these cases have these features and aren’t fitting objects of righteous indignation.

I don’t think Lemos’s argument here provides good reason to favor his explanation over my explanation because I think there is a problem with using this thought experiment as a litmus test for
the different explanations. Thought experiments like this are often offered as challenges to desire satisfaction theories of well-being: the desires appear ‘defective’, such that the satisfaction of them doesn’t seem to make the person who has them better off. Since these cases are viewed as challenges for desire theories of well-being in general, it’s unsurprising that we find them odd or curious. And because of this, I think it’s misguided to rely on them to adjudicate between Lemos’s explanation and my own.

Lemos’s second anticipated objection is that of ontological parsimony: the explanation that (1) is a fitting object of indignation because A’s being pleased in good for A and contrary to what A deserves requires two evaluative notions—final welfare value and final value simpliciter—whereas his explanation only posits one, i.e. final value simpliciter (1998: 335). And, since his account is theoretically simpler, there’s reason to prefer it.

I don’t think that this is a convincing reason to prefer Lemos’s account for two reasons. In making this claim, Lemos adopts a Moorean conception of well-being: the only things that are valuable are finally valuable simpliciter, and a person’s good is simply having those things in their life (Hurka 2015: 9). For those who agree with these types of accounts of well-being, this may indeed provide reason to favor his explanation. But my explanation carries no worrisome baggage for those who already think of well-being as a distinct kind of value. So, at least, those who antecedently reject a Moorean picture of well-being should see Lemos’s challenge of theoretical simplicity as no threat to my argument.

But, of course, Lemos’s parsimony point may be appealing to those who are undecided or uncommitted regarding the ontological status of prudential value. Unfortunately, settling the monism/pluralism debate about value categories is an ambitious project that is outside the scope of this paper, but I think that there are reasons to be more permissive. One reason to think so is that

---

22 See, e.g., Heathwood (2005) for discussion of these.
there seem to be cases where moral value and well-being conflict: it may improve one’s well-being to
to cut in line (less time waiting at the DMV!), cheat on one’s spouse (supposing that there are no
negative consequences like getting caught or feeling guilty), not helping a colleague move (you’re
tired and they’re just using you for your truck), etc., but there seem to be moral reasons not to do
these things (e.g., fairness, honoring promises). This is easily explained by positing more than one
evaluative concept. Using both concepts also gives us a plausible reading on Ross’s Two Worlds
cases (in which two worlds have the same amount of pleasure and differ only in that one is filled
with virtuous people and the other with vicious people): while both worlds contain the same amount
of prudential value, the world filled with vicious people is morally worse (perhaps worse simpliciter?)
than the one with virtuous people. Thus, there’s prima facie explanatory power for positing different
kinds of value, and so there’s at least some reason to favor a more permissive ontology.

V. A Reductionist Challenge

Suppose now that my analysis is correct and that Lemos’s argument establishes Strong Invariabilism
for final welfare value. This might be leveraged to argue for Strong Invariabilism for final value
simpliciter. In this concluding section, I show how it might be done and offer an argument for why
this approach fails.

While I have been treating final value simpliciter and final welfare value as distinct (and,
indeed, argued against a reduction of welfare to value simpliciter), one might argue that we should
analyze the former in terms of the latter. As Heathwood points out, skepticism about the existence
of welfare value seems less common than skepticism of value simpliciter, and “this differential in
perceived dubiousness is one motivation for such a reduction” (2015: 139). The following is a proposed reduction of value simpliciter to welfare:

Welfare Only (WO): a state of affairs $\Phi$ has a final value of $V$ just in case $\Phi$ contains a balance of welfare at level $V$.

One can use WO in conjunction with my amended conclusion of Lemos’s argument to argue that Strong Invariabilism is true for final value simpliciter. If Strong Invariabilism is true for final well-being and final value simpliciter reduces to final well-being, then Strong Invariabilism is true about final value simpliciter.

A full evaluation of WO is outside the scope of this paper. But I think a good reason to doubt WO is that it is inconsistent with the following plausible principle:

Weak Prioritarian Principle (WPP): For an increase in well-being $D$ and for two people $A$ and $B$, where (i) $A$ is much worse off than $B$ and (ii) there is no other morally relevant difference between $A$ and $B$, $A$’s receiving $D$ is at least slightly finally better simpliciter than $B$’s receiving $D$.

Suppose Alex has a low level of well-being (20) and Betty has a very high level of well-being (2,000). By WPP, it would be at least slightly finally better simpliciter to give a welfare benefit of 10 to Alex than to Betty if we could only give the benefit to one of them. A world in which we benefited Alex (Alex at 30 and Betty at 2,000) would be at least a slightly more finally valuable world than one in which we gave the benefit to Betty (Alex at 20 and Betty at 2,010). This strikes me—and philosophers such as Carter (2011)—as plausible.24

---

23 For those skeptical about the notion of value simpliciter, see Geach (1956); Foot (1985); Thomson (2008); Kraut (2011).
24 “But increasing the happiness of a barely happy person seems more urgent than increasing the happiness of an extremely happy person. An obvious explanation is that more value is obtained from the former change than from the latter” (Carter 2011: 394).
Suppose, on the other hand, that WPP is false. If WPP is false, then it would be no more valuable *simpliciter* that a person with a very low level of well-being—e.g., a person with a life barely worth living—receive a benefit in well-being than a person with a very high level of well-being receiving a benefit of the same magnitude. This seems wildly implausible. So, since WPP captures the intuitively correct answer in the Alex/Betty case and denying it seems wildly implausible, there is at least some support for WPP.

WPP and WO are incompatible. WPP entails that the amount of final value *simpliciter* can differ even when the total increase in well-being is the same, while WO holds that the same increase in well-being necessitates the same increase in value *simpliciter*. Because they are incompatible and the above considerations offer support for WPP, there is some reason to doubt WO. And, if WO is false, then my concession that Lemos’s argument establishes Strong Invariabilism for final well-being doesn’t require that I concede Strong Invariabilism is true for final value *simpliciter*.

In conclusion, I have argued that, while one of the primary arguments for Strong Invariabilism about final value *simpliciter* fails to establish Strong Invariabilism for final value *simpliciter*, it plausibly establishes Strong Invariabilism for final welfare value. I have identified how one might attempt to use this to argue for Strong Invariabilism for final value *simpliciter*—via WO—and have shown that WO is incompatible with what I’ve briefly argued to be a plausible principle (WPP).

References:


---

25 Acknowledgments omitted for blind review.