

***Adversus “Adversus Homo Economicus”:***  
**Critique of the “Critique of Lester’s Account of Instrumental Rationality”**

J. C. Lester

***Abstract***

*This essay goes through Frederick 2015 (the critique) in some detail, responding to the various paraphrases and criticisms therein. It is argued that in each case the critique is mistaken about what Lester 2012 (Escape from Leviathan: EfL) says, or about what the critique presents as a sound criticism, or both. Introduction: the three problems with the critique and the philosophical problem that EfL is attempting to solve. “Abstract”: the critique’s confusion about EfL’s aprioristic theory of instrumental rationality. There are then detailed replies (too many and too diverse to summarise) to quotations from the critique’s confused interpretations and criticisms under the following headings (quoted from the critique): “2. Instrumental Rationality”; “3. Weakness of Will”; “4. Desires and Values”; “5. Free Will”; “6. Self-Interest”; “7. Maximisation”; “8. Morals”. Finally, in “9. Conclusion”, the philosophical problem for economics is briefly restated. The philosophical interpretation of homo economicus in EfL (as with its overall philosophical theory of libertarianism) has yet to be given adequate critical consideration.*

**Introduction**

There are three major and general problems with Frederick 2015 (the critique) that occur repeatedly throughout. (1) Not showing any grasp of the philosophical problem being tackled and how the philosophical theory is supposed to solve it. (2) Paraphrasing with great inaccuracy, instead of quoting, and then replying to the inaccurate paraphrases. (3) Repeating the same, or very similar, points. There are various ways of dealing with these problems, including the following three general approaches. One way would be to dismiss the whole essay as substantially irrelevant unless and until it is rewritten to deal with these points; but then such rewriting seems unlikely to happen.<sup>1</sup> Another way is to deal with each type of criticism only the first time that it occurs in the critique; but this could make for a very short reply that might be perceived as not doing justice to the detail in that long essay, possibly because of a perceived failure to be able to respond to those details. A third way is to go through it all fairly thoroughly; but this must result in much repetition of similar points that might try the patience of the reader. On balance, the third way seems to be the least of the three evils and so that has been chosen.

To put things into context, then, it might help to start by outlining what philosophical problem *EfL* chapter 2 is trying to solve and how it tries to solve it; something that the critique fails to do, just as Frederick 2013 (the liberty critique) also fails to do.

To quote from the opening paragraph of that chapter:

The standard, modern, economic assumption of instrumental rationality holds people to be self-interested utility-maximizers. Economists usually intend this to mean egoistic preferences and perfect calculation over time. Insofar as people think this to be unrealistic, it throws doubt on the (generally pro-market) conclusions of the economists. Here we give the assumption of self-interested utility-maximization an aprioristic interpretation that may help to reconcile standard economics with Austrian School aprioristic economics. This *a priori* sense does not imply egoism and is not trivial. Economics requires this sense as a core assumption in order to link its results more convincingly to liberty, welfare, and anarchy. This book’s conceptions of liberty, welfare, and anarchy also presuppose some such account. (11)

---

<sup>1</sup> Especially since the author has now died.

This can be explained more generally. Assume a central aprioristic<sup>2</sup> theory of instrumental rationality, as an interpretation of ‘economic man’. People are thereby—other things being equal—better, albeit still fallible, judges of their own liberty (deciding which things are free personal choices) and welfare (deciding which things makes them better off). And their empirical choices will tend to reveal their judgements. So, if we disagree with their choices—whether those choices are purely personal or initiate impositions on other people—then it usually makes sense to reason with them, at least in the first place. Without the assumption of such core rationality, it is common to view people as sufficiently irrational as to make at least some authoritarian paternalism a default assumption (although this appears to leave any actual authoritarian paternalists as equally irrational, and that is before considering the limits of knowledge with respect to ruling other people efficiently, and the temptations of such power). People are thereby often supposed to smoke or take ‘drugs’, drink ‘too much’ alcohol, eat ‘too much’ of the ‘wrong’ foods, not exercise ‘enough’, ‘bigotedly’ discriminate against certain people or practices, and generally not to know what they ‘rationally’ ought to do (whether prudentially or morally), or to know it but fail to do it anyway—all because they are ‘incurably irrational’. Hence, they often need to be ruled by proactive coercion, both for their own sakes and the sakes of the other people with whom they interact. Belief in this view is one important reason that the popularly-elected oligarchs of modern states pay only lip service to both liberty<sup>3</sup> and democracy<sup>4</sup> in order to help to legitimise their rule. In reality, both liberty and democracy conflict greatly with their own preferred, and jealously preserved, oligarchical rule (as well as conflicting with each other). Consequently, if one is defending individual liberty both as an end in itself and as welfare-enhancing, then the hypothesis that people have no core instrumental rationality is a problem, and the hypothesis that they do is a solution. Moreover, many of the other ascriptions of ‘irrationality’ do not appear to survive philosophical scrutiny.<sup>5</sup> None of this is to suggest that people are anywhere near being perfectly prudent, or that they do not make many mistakes.

This can be given more context still. As the subtitle of the hardback of *EfL*—and now subtitle of the paperback—states, the book is about “Rationality, Liberty, Welfare, and Anarchy Reconciled”: their systematic congruence being an extreme form of the, usually implicit, “classical-liberal compatibility thesis”. That reconciliation is about providing and defending philosophical theories of economic rationality, interpersonal liberty, want-satisfaction welfare, and private-property anarchy. This is done in order to make better sense of their apparent consistency in terms of the social scientific evidence; from economics, in particular, but also politics, history, sociology, psychology, etc.. The often perceived ‘improbable coincidence’ of their real compatibility has caused some critics erroneously to suppose that *EfL* is providing what are mere stipulative *definitions* that have simply been engineered in order to fit together. But *EfL* clearly explains how each one is a genuine *theory*, and how it could be false or clash with the others. However, there are both philosophical and, to a lesser extent, empirical explanations that they are not false and do not clash.

This essay will now go through the critical responses in the critique in the order in which they appear, dealing with most of the errors therein.

## “Abstract”

The “Abstract” of the critique includes the statement that *EfL*

defends two hypotheses: that instrumental rationality requires agents to maximise the satisfaction of their wants and that all agents actually meet this requirement. (1)

---

<sup>2</sup> Or, at least, at the more *a priori* end of the spectrum with *a posteriori*—if that distinction is not ultimately defensible as completely separable. Also, in accordance with critical rationalism, “*a priori*” is here intended conjecturally rather than as ‘justified’, proven, or apodictic.

<sup>3</sup> ‘Liberty’ is here understood in the orthodox classical liberal and libertarian sense of, in concrete practice, the absence—failing which, minimisation—of ‘invasions’ or ‘aggressions’ on people and their non-invasively-acquired property (where this must ultimately, but more controversially, be derivable from an abstract theory of liberty such as ‘the absence of interpersonal initiated imposed costs’; see Lester 2012, 2014, and 2016).

<sup>4</sup> ‘Democracy’ is here understood in the classical sense of direct rule by the people in some manner; such as by plebiscite or sortition.

<sup>5</sup> See, especially, Percival 2012.

This is mistaken. The utility-maximisation aspect is explained in *EfL* as follows:

4) Utility-maximization: What one has the strongest desire to do, one has the strongest utility at the thought of doing. Agents aim at the goal the thought of which gives them the greatest utility, or least disutility, at the time of aiming at it .... (14)

There is no statement or implication that agents, are trying to, “maximise the satisfaction of their wants” over time or that “all agents actually meet this requirement”. An inaccurate paraphrase is used, and later criticised, instead of an accurate quotation—as also occurs throughout the liberty critique.

## “2. Instrumental Rationality”

The *a priori* theory of economics’ ‘instrumental rationality’ (self-interested utility-maximisation) that *EfL* is defending can be expressed thus: everyone has self-perceived interests (i.e., desires/wants/valued outcomes), whether these are egoistic or altruistic, and he always attempts to act on the interest that maximises his utility (i.e., gives him the greatest desire-/want-/value-satisfaction) at the time of the attempted action in the circumstances as he, conjecturally, perceives them to be.

The critique tells us that

A minimal conception of instrumental rationality is this: an action is instrumentally rational if and only if it is a suitable means to the achievement of the agent’s ends (2)

A *more* minimal conception (in the sense that it is, more or less, *a priori*) of instrumental rationality is this: an action is instrumentally rational if and only if it is *intended to be* a suitable means to the achievement of the agent’s *current subjective* ends. And that is, roughly, the conception that is used and defended in *EfL*.

But the critique goes on to assert that

As it stands, this statement may be interpreted objectively, subjectively, or part one way, part another. (2)

“As it stands”, the critique’s statement is ostensibly only objective: “it *is* a suitable means to the achievement of the agent’s [actual/objective] ends”.

The critique continues,

Lester undertakes to defend a thesis, which he labels ‘instrumental rationality,’ according to which every agent seeks to maximise the satisfaction of his wants over time (2012, 13-16). (2)

As shown, the cited pages explicitly reject that. There is no such hypothesis in *EfL* (if there is, then why not quote it?). Hence what immediately follows on from this in the critique is irrelevant.

The critique goes on,

The second half of Lester’s *a priori* thesis of instrumental rationality is a hypothesis about agency, namely,

(LA) all actions are instrumentally rational.

Only by intention under the perceived circumstances. But the critique immediately rejects this:

A qualification needs to be added. ... we should say that the agent will act to satisfy his strongest desire provided the action is possible for him and he thinks that it is. (3)

No: we *always* do *attempt* to satisfy our strongest desire in the circumstances as we conjecturally perceive them to be. The action need not be possible.

### “3. Weakness of Will”

As *EfL* says, nothing is here meant to deny the existence of ‘weakness of will’ in the sense of ‘lack of determination or willpower’. For instance, Peter might be able to hold his hand in painfully cold water significantly longer than Paul (assuming, *arguendo*, very similar subjective experiences of pain). That some people are ‘strong-willed’ and others are ‘weak-willed’ in terms of ‘determination’ or ‘willpower’ is not the same sense as occurs in discussions of supposed ‘*akrasia*’: meaning, ‘not acting in the way that one believes to be the best’. More centrally, *EfL* holds that much confusion is avoided here by distinguishing levels of desire/want/value. One may have a first-level (or immediate) desire/want/value in favour of X, but a second-level (or meta) desire/want/value that is against the first-level one. And the stronger of the two levels necessarily prevails. When the first-level is stronger, and so prevails, that is mistaken for *akrasia*. (In principle there could be a third level, and so on, but such complications do not need to be discussed here.)

The critique asserts that *EfL*’s exposition of how weakness of will (i.e., *akrasia*) can be explained away “is not entirely clear” (4). But there is no quotation to illustrate this alleged lack of clarity. Therefore, as usual, an attempted paraphrase by the critique follows instead (here in three parts):

- (i) of all the agent’s first-level desires for action in the circumstances, the agent’s strongest desire is the desire to perform the action he does perform;

Wrong: he *attempts* to perform the action consonant with his strongest first-level desire in the *conjecturally perceived* circumstances—*unless* there is a relevant stronger second-level desire (such as a desire to resist that desire).

- (ii) that action has consequences, or aspects, which the agent desires not to be, and this first-level desire is greater than any second-level desire he has to retain the desire to perform the action;

Where this says “to retain the desire”, this might be an error for ‘to refrain from the desire’. Otherwise, how does it make sense? But again, in any case, so much for the critique’s attempt at being “entirely clear”.

- (iii) as a consequence of (ii), of all the agent’s second-level desires about the first-level desires in play in the circumstances, the agent’s strongest second-level desire is the desire not to have the desire to perform the action.

Whether and how and this is a “consequence of (ii)” is not clear. In any case, the first-level desire to perform the action is, at that time, *stronger* than the second-level “desire not to have [or to refrain from acting on] the desire to perform the action”.

To supposedly illustrate its three parts, the critique then quotes the examples of the smoker and the chocolate-eater in *EfL* and asserts that “this alternative account of apparent cases of weakness of will is not convincing” (4). This is because

- Contrary to (i), the smoker who suffers weakness of will fulfils his desire to continue smoking despite the fact that he thinks he would be better off if he did not.

This is simply to restate the common sense, but paradoxical, view of weakness of will (*akrasia*). The two-levels of desire explain away the paradox with *akrasia* that the critique apparently fails to see, because for the critique having a greatest “desire” for X and believing that ~X makes one “better off” are held to be incommensurable phenomena. But “desires” (or what is wanted) and what is thought “better” (or what is valued) are often translatable into each other: in some obvious sense, someone must value what he desires/wants and must desire/want (in some way) what he values. So, it appears to be inconsistent to say ‘I *most* desire (or want) X but I think best (or *most* value) ~X’, hence the paradox of *akrasia*. We must want what we genuinely think, at the time, is overall best under the perceived circumstances.

The critique continues:

Similarly, the woman suffering weakness of will does not feel that life without satisfying her desire for chocolate is too miserable to forego satisfying that desire. On the contrary, she feels that life without satisfying that desire would be better than life in which the desire is satisfied. (4)

Again, this merely reasserts the paradoxical common-sense view whereby we do not do what we supposedly think it better to do.

The critique asserts that

These seem to be descriptions of ordinary facts of life and ones which many smokers and chocolate-eaters offer as a description of their situations.

But how can these be *true* “descriptions of ordinary facts of life”? This is analogous with hearing evidence that the Earth, other planets, and stars revolve around the sun (heliocentrism), and then saying in response to that evidence that the sun, etc., going around the Earth (geocentrism) is one of the “descriptions of ordinary facts of life”.<sup>6</sup> This ignores the problem and the proffered solution and simply restates the common-sense view.

The critique goes on to say that

Often people make such assertions, or have such thoughts, as a prelude to taking steps to prevent themselves from smoking or from eating chocolate. Such steps may involve other people who can help to restrain them from actions they strongly desire to perform but which they think it is better not to perform. (4)

There is a coherent distinction between actions people “strongly desire” and those that they “think it is better not to perform”. But can this distinction between “desire” and what is “better” be coherently made in the context of *akrasia*? For that would be saying that what we ‘*overall most* desire to have’ need not be what we think to be ‘*overall best* to have’. The appetitive connotations of “desire” are not relevant here. ‘Most desired’ means only ‘most wanted’—for whatever reasons, including prudence and morality. So, one is supposed to hold both that, 1) all things currently considered, at this moment, I most want X, and 2) all things currently considered, at this moment, I think it best that I do not have X. But what you ‘most want to have’ you must, in some obvious sense, think it ‘best to have’, and *vice versa*. Admittedly, it is not possible to deduce a clear contradiction (A & ~A) if there is an insistence on a distinction between what is ‘most desired/wanted’ and what is thought ‘best’. But there is an intuitive inconsistency—hence the problem with *akrasia*—and so the alleged distinction is not at all clear.

That said, there are two relevant ways of interpreting the above quotation. One is binding one’s future self and the other is binding one’s current self. There is no weakness-of-will paradox in binding one’s future self. People can consistently ask other people to “help to restrain them from actions they [will most] strongly desire to perform [later at T1] but which they think [now at T0] it is better not to perform.”<sup>7</sup> And the distance in time between T0 and T1 can be a mere moment. It is only all this happening at the same time that is inconsistent. And, as explained, it is not clear that the critique can avoid this inconsistency by trying to distinguish a “desire” from what one genuinely thinks is “better”. For what is inconsistent is, at time T0, both most desiring (thereby thinking it overall best in some sense) to perform some action and thinking it overall best (thereby most desiring in some sense) to be restrained from performing it.

The critique then makes an interesting and related suggestion:

Contrary to (iii), the weak-willed agent might not desire to be rid of the desire to perform the action he performs. For example, the woman who eats chocolate even though she most wants to stop eating it,

---

<sup>6</sup> Neither theory is true, but heliocentrism appears to have more verisimilitude (despite the logical problems with that idea).

<sup>7</sup> The possibility that one could “contract into a penalty system” is mentioned in *EfL* (25).

Here, incidentally, the critique slips into clear inconsistency. How can one willingly do what one “most wants” not to do? The critique continues that she

may desire to retain her desire to eat chocolate; and her desire to retain her desire to eat chocolate may be stronger than any desire she has to be rid of the desire to eat chocolate. For, although she would rather not *eat* chocolate, she may value very highly her *desire* to eat chocolate, perhaps because she values having that desire for its own sake, or perhaps because she values having that desire frustrated, either intrinsically or as evidence of her willpower (or for some other reason). Therefore, while her first-level desire not to be fat conflicts with her first-level desire to eat chocolate, it need not conflict with her second-level desire to retain her desire to eat chocolate. (5)

It is true that one can desire to keep a desire that is problematic in some respects. To respond relevantly, we first have to make the second-level desire more precise: the desire (or thinking it overall ‘better’) that the first-level desire continues but is somehow “frustrated”. In the previous case, the stronger desire to have or do X overrides the weaker desire to not have—or to resist—the desire to have or do X. But now there is no direct clash of desires. For the desire (or thinking it overall ‘better’) to have some desire but have it “frustrated” is merely an idle wish unless we can act on it. However, if we can act on it (e.g., pressing this button will somehow cause the frustration of the first-level desire), then we are back with a direct clash (the agent will only press the button if his second-level desire to have but frustrate his first-level desire is stronger than that first-level desire). Therefore, the second-level ‘desire to have a desire but have it “frustrated”’ fits the two-level analysis just as well as the ‘desire not to have, or to resist, a desire’ (and doubtless there are other types of second-level desire that also fit). The apparent paradox of *akrasia* has still been explained away. So, the critique remains confused to conclude that “The phenomenon of weakness of will refutes (LA). (5)”.

#### “4. Desires and Values”

It is not unusual to make a linguistic distinction between ‘desires’ and ‘values’. ‘Desires’ are often thought of as more basic and more appetitive, while ‘values’ are often thought of as higher and more reasoned. Thus, desires and values can appear to be quite different, and even incommensurable, things. But *EfL* observes that to desire something must in some sense always be to value it (if only as a way of satisfying that desire) and to value something must in some sense always be to desire it (if only as a way of achieving that value). Thus, no essential difference, or incommensurability, need impede the instrumental rationality that is being defended.

The critique quotes Watson 1975 on the supposed distinction between reason and desire. It is then asserted that

In contrast, Lester (2012, 28-31) thinks there is only a single source of motivation, of reasons for action. We desire something if and only if we value it; though to say that an agent values a thing is not to spell out the nature of his desires about it. (5)

Why does the critique repeatedly resort to dubious paraphrase instead of relevant quotation? It is less clear and less useful to have to reply to such gratuitous imprecision. The critique continues,

It is, Lester claims, absurd to say that someone is inclined to do something yet does not have even a *prima-facie* reason to do it. (5)

Or as *EfL* *actually* says, and in a little more context,

Here ‘reason’ is again apparently being used in the sense of a value that we have arrived at verbally. That, at least, is the only real sense I can make of Watson’s distinction. This may be a consistent way of talking about things, but it is confusing. For in plain language, it is absurd to say that someone is inclined to do something yet does not have even a *prima facie* reason to do it. If he can see no reason *whatever* to do it, then he can hardly desire it. Desire is at least a *prima facie* reason to do a thing. In Watson’s examples the benefits of acting on such desires are really seen as hugely outweighed by the costs. (31)

The critique responds to the examples in Watson 1975 as follows:

It seems to me that this argument about whether the woman and the man, in Watson's examples, *value* as well as desire to perform the actions in question, is a dispute over linguistic nuance. The same goes for whether we should say that each of those agents has a *pro-tanto* reason for, or a consideration in favour of, the actions. (5)

The idea that philosophy (*qua* philosophy) is ever arguing about "linguistic" usage is false and even philistine. To have a desire for something is *ipso facto* to value that thing and to have a reason to have it, at least in that respect and to that extent. To deny that there is any real value or reason here is to deny the facts.

The critique continues,

But there is a real distinction behind Watson's more restrictive linguistic proposal. The woman thinks that drowning her child would not be an objectively good or valuable action, even though she desires to do it and, therefore, in some sense, values it.

She does not merely think that "drowning her child would not be an objectively good or valuable action". She most likely thinks it to be clearly immoral, and thereby categorically undesirable/unacceptable (although *EfL* does not discuss morals at this point). And this "real distinction" in a type of desire is thereby not being denied but, rather, explained (whether, and in what sense, that moral view is held "objectively" is a separate matter). As discussed later in *EfL*, and also defended later here, what we genuinely hold to be immoral we thereby desire to avoid at all costs. Therefore, her desire to avoid an act felt to be immoral necessarily trumps her weaker desire to do that act. There is no need to abandon the realm of desires or wants. This matters because of the core rationality thesis that is being defended.

The critique then asserts that

The man may, in some sense, value sexual activities, given that he desires them; but he thinks that it would be objectively bad or wrong to satisfy those desires. If, with Lester, we want to say that people always value what they desire, then we can say that the woman and the man value something they think is not (objectively) valuable.

Or as *EfL* would interpret it, they feel it to be immoral and hence categorically undesirable/unacceptable; and so, *ipso facto*, they will not choose to act in those ways. It is not necessary to posit the realm of the "(objectively) valuable".

The critique says that "Lester recognises (2012, note 46, 209), that this is the point behind Watson's proposal" but

he misstates what is at issue when he says that Watson needs an account of objective values. That is not quite right. To make sense of the man and the woman, what Watson needs is to attribute to them a belief in objective values; but he does not need to endorse that belief himself. (6)

*EfL* does not say or imply that Watson does or need "endorse" a belief in objective values. It only says that "perhaps" he needs "an account of objective values". Because without an *intelligible* "account of objective values" we cannot even "make *sense* of the man and the woman ... [having] ... a belief in objective values".

The critique continues:

Lester seems reluctant to accept that there are objective values,

Speaking literally, values are inherently subjective: they are someone's feelings of value.

and because he mistakenly thinks that acceptance of such is required to give the account of the woman and the man that I just gave,

Not “acceptance”, merely intelligibility.

he offers a different account, one which invokes, again, the distinction between levels of desire. (6)

That idea is merely mentioned in passing on these issues. The main point in *EfL* here is, rather, the general one that in “Watson’s examples the benefits of acting on such desires are really seen as hugely outweighed by the costs” (31). But the critique ignores this given explanation and launches into its own “levels of desire” account (6) that is simply not in *EfL* on these issues. In any case, moral values/feelings/desires can also be at the first (immediate) level. What else is wrong with the critique’s objections to a “level analysis” has already been explained.

The critique claims

We have seen that, on Watson’s view, an agent has desires and values, and only the latter give (legitimate) reasons for acting, while on Lester’s view an agent simply has desires, which we may also call ‘values.’

We have seen that Watson 1975 and the critique are confused. We can only act on the basis of what we desire/want/value. That does not in any way entail or imply that all desires/wants/values are equivalent in terms of prudence, or morals, or importance, or critical preference.

The critique’s confusion continues: “Lester seems to see all desires as appetites (2012, 16)”. But the fact that *EfL* happens to use “appetite” in a broad sense on that particular page does not mean that a narrow sense is being denied. When “Watson distinguishes appetitive or passionate desires from those which are the products of culture or habituation”, *EfL* can happily allow this, as it in no way conflicts with the general instrumental rationality that *EfL* is defending. When the critique says that the “latter rank as desires rather than values for Watson because, being merely inherited, they are not the products of the agent’s rational judgement [i.e., critical preference] (1975, 214-15)”, this overlooks that all sincere and literal values must be felt (valued) and thereby, in some sense, imply relevant desires. All this seems to be another case of the critique’s insisting on keeping conventional distinctions—which *EfL* innocuously analyses in terms of general desires or wants in order to defend instrumental rationality—rather than showing that *EfL*’s analysis is genuinely flawed.

The critique then suggests that we “distinguish” the following:

- (1) an agent’s felt desires;
- (2) things an agent thinks are valuable;
- (3) things an agent thinks are valuable for him;
- (4) an agent’s goals. (6)

All of these things can be colloquially distinguished. But that is irrelevant to the problem that *EfL* is addressing: whether instrumental rationality can be defended by explaining all such things in terms of an agent’s attempting to maximise his desire/want satisfaction (in the perceived circumstances, at the time of the attempt). This is not to imply that they are all illusory distinctions that are completely reducible to the same thing, or that they are all equivalent in terms of morals, or prudence, or importance, etc. And, clearly, they can be thus explained. “(1) an agent’s felt desires”: these are unproblematic. “(2) things an agent thinks are valuable”: an agent must in some sense still desire these, if only desire that they exist. “(3): things an agent thinks are valuable for him”: he must in some sense desire these for himself. “(4) an agent’s goals”: however these are construed, they cannot be his genuine goals unless he desires to realise them in some way. The critique asserts that theorists “have tended to conflate these” four things. But to point out the necessary role of desire- (or want-) satisfaction in them all is not to “conflate” them. By analogy, to point out the necessary role of physical competition in different sports is not to “conflate” those different sports either.

The critique asserts,

I may value something that I do not desire, and *vice versa*. For example, if I am not hungry and lack a desire to eat, I may nevertheless eat because I value eating for social or nutritional reasons;

This all makes good colloquial sense, of course. However, to eat for “social or nutritional reasons” is to act on the desire to eat for those reasons. Desire cannot be eliminated. And that is all that our *homo economicus* requires.

or I may go to work despite feeling no desire to do so, because I think going to work is valuable for me.

In other words, there is a desire to work but for reasons other than the work itself. The work is a means and not an end.

Alternatively, I may be hungry and feel a strong desire to eat, but not eat because I am on a diet and think that eating is not valuable for me;

In other words, this is to desire the consequences of not eating more than giving in to a “strong desire to eat”.

or I may yearn to get back to work but remain at home because I am recovering from illness and I value a swift and full recovery.

In other words, this is to desire a “swift and full recovery” more than the immediate desire “to get back to work”.

However, I may also both desire and value something, as when I eat because I have a healthy appetite.

In other words, this is both to desire to eat and to desire that desire because of the good consequences.

So far I have spoken of what is valuable for the agent. But all except the most self-centred of people think that many things have value even though those things are not valuable for the person himself. (6-7)

Things that are “not valuable for the person himself” in a narrow self-interested sense, may still be “valuable for the person himself” in the altruistic sense that he values (and thereby desires) certain goods for other people (or animals, or the environment, etc.).

For example, someone might think that classical music is valuable for some people, and thus valuable, even though he thinks (correctly, perhaps) that it is not valuable for him.

This is ambiguous. Either someone has an altruistic value (and thereby desire) that other people can consume classical music if they value (and thereby desire) it, or he is merely aware that some people value (and thereby desire) classical music. Whichever is meant, this is not inconsistent with the given account of how values have a desire aspect.

This person may have no felt desire for classical music, despite valuing it.

In other words, he has “no felt desire for classical music” in his own life, but he values it and thereby has a “felt desire for classical music” in the lives of people who do personally value it (and thereby personally desire it).

The critique then moves on to “(4) an agent’s goals”:

An agent’s goal may be to perform an action of a particular type even though performing an action of that type is neither desired nor valued by him. (7)

Not clearly appetitively “desired” or deliberately “valued”, perhaps. But such narrow interpretations are not ‘true essences’ that can refute broader interpretations.

For example, when I come downstairs in the mornings, while I am waiting for the kettle to boil, I pour and drink a glass of orange juice. I do this as a matter of routine, without thinking about it.

Or, more precisely, without deliberation or significant self-awareness. Some form of thought is required.

Sometimes I only know I have drunk the orange juice because I can see the used glass. On some such occasions I will not have had a felt desire for the orange juice nor will I have thought that drinking it was valuable for me or valuable for anyone else. I just drank it out of habit.

Let it be granted that he “just drank it out of habit”. We can only engage in a habit because of some sort of a desire to do so. Such actions might not require significant deliberation or significant self-awareness, but there must be some conscious desire or we would be, temporarily at least, like unconscious automata (forgetting how we felt at the time about some habitual action seems far more plausible than that we were literally not conscious at the time). And what we desire we must thereby, in some obvious sense, also value—if not value in any second-level sense that we might colloquially reserve for talk about ‘values’ and what is “valuable”.

The critique continues:

the action was intentional, that is, goal-directed: I was in control of what I was doing and of whether I was doing it (I could stop if I wanted); it was something that I aimed to do ... and if, as I was about to do it, I was asked what I was going to do, I could truly have said that I intend to drink the juice. Thus, we can have goals that we neither value nor desire

“I could stop if I wanted” implies that he wanted to act as he was doing. But to want something must be to have *some* sort of desire to have it and in *some* way to value having it. So, it looks like another fairly plain, and quite perverse, inconsistency to hold that one’s “action was intentional, that is, goal-directed” but in *no sense* desired or valued. And it is only very general senses of ‘want’, ‘desire’, and ‘value’ that apriorist instrumental rationality requires.

The critique says that “in practice, the term ‘desire’ or ‘want’ is extended to such cases”:

a person who knows my habit and who sees me go, absent-mindedly, to the fridge in the morning, may say that I want to drink some orange juice, or that I desire some orange juice. But, in this sense, to say what I ‘want’ or ‘desire’ to do is just to say what I *aim* to do.

But how can one consistently “*aim* to do” something without a plain ‘want’ or ‘desire’ to do it? These terms do not seem to be “extended” in any metaphorical or “attenuated” (7) sense. Even if what one has an “*aim* to do” is what is habitual without any thought whatsoever as to the content of the habit (if that is possible), still that habitual “*aim*” itself must be what one does ‘want’ or ‘desire’ to do or achieve. The critique is apparently restricting the meaning of ‘desire’ and ‘want’ to something directly appetitive. This seems to be essentialist language-policing. In this case, it both disallows an innocuous, and useful, account of instrumental rationality and entails, instead, an intuitively incoherent way of describing actions.

## “5. Free Will”

In *EfL*, an agent acts out of ‘free will’ when his body moves as he chooses to make it move. This Hobbesian sense of ‘free will’ is compatible with someone’s threatening to shoot him if he does not do what is demanded, and so he does it. *EfL* supposes that all acts of free will involve the agent’s trying to achieve what he most desires/wants/values (i.e., gets the most utility/satisfaction at the thought of doing) at the time, in the circumstance as he conjecturally perceives them to be. *EfL* also supposes that an agent’s free will cannot escape physical determinism (he chooses but as he must choose) without falling into the randomness of indeterminism. However, this ‘compatibilist’ assumption is not necessary for instrumental rationality.

In the critique we are told that *EfL* “employs a passive conception of agency” because the “agent always performs the action that he thinks will most satisfy his wants” (7). It is not yet clear how this is supposed to be “passive”. Also, it is only about *attempting* to perform the action that *does* “most satisfy his wants” *at that time* (although this might relate to the future, or to altruistic concerns).

According to the critique, this is “not a conception of agency at all” because

the supposed action is brought about by the agent’s motivational factors: the supposed agent does not act, but is rather the passive recipient of impulses which propel him hither and thither. The agent’s body moves in response to desires or values, but the agent does not act. (8)

*EfL*’s account is fully compatible with the agent’s reasoning carefully about his possible courses of action to select which one is the most preferred. Of course, he has first to find the idea of reasoning carefully to be his most preferred option. But it is not clear how an agent can possibly avoid ending up with attempting to do what he most desires or wants to do (at the time in the perceived circumstances). To not do that would appear to mean ‘acting’ without motivation (or, at least, with a lesser motivation).

The critique continues:

Curiously, Lester contrasts his view of the agent, as following his consciously felt, self-perceived interests or desires, with a view of the agent as an unconscious automaton without a spontaneous will of his own (2012, 14). But how can someone who is doomed always to follow his strongest desire be said to have a spontaneous will of his own? If Lester’s agent can be distinguished from an unconscious automaton, it is only because he is a conscious automaton.

How can we be “doomed” to being able to, try to, do what we most want to do? Especially as that will include as much reasoning, ambition, morality, self-control, etc., as we want. The critique asserts that “if actions are always dictated by the strongest desire, the agent has no choice”. Again, “dictated” is an odd word to choose. So, what is supposed to be a more attractive alternative?

The critique tells us that it is a sense of free will that is

incompatible with determinism ... incompatible with that agent’s actions being determined by prior circumstances, whether or not those circumstances include desires or valuations. (8)

One thing to mention first is that in Lester 2017 (25-26) there is an extremely brief and highly speculative account of how intellectual activity might escape physical determinism without falling into mere randomness. That said, it does not appear to be relevant here which account of free will is right: compatibilist or incompatibilist. It is not about causality but the intelligibility of the account. It does not make sense that an agent finds action X to be the most desired/wanted/valued at time T, but intentionally does Y at time T instead. This is like the alleged existence of *akrasia*, but more clearly inconsistent.

The critique quotes *EfL* where it says “‘The school of thought that demands a kind of free will that escapes both determinism and mere randomness has never given an intelligible account of a third option’ (2012, 21)”. The critique calls this “the infamous ‘chance objection’ to free will: if my actions are not determined, they are a matter of chance; but if they are a matter of chance, they are outside of my control”. But “chance” does not seem quite appropriate, as it appears compatible with determinism that some things happen by “chance” in the sense that they are not intended or predictable. The critique says that this is “confusing an undetermined act with a random event” and asserts instead that an “act, specifically, an act of will, is something that is inherently under the agent’s control and that is therefore undetermined”. How this escapes the compatibilist account of free will without falling into randomness is left mysterious. Perhaps there is an explanation in the references the critique gives. However, as Lester 2017 does give an attempted explanation, and as this is, in any case, irrelevant to the *a priori* argument just given, those references won’t be pursued here.

The critique then offers the account that

many substantial goals that agents pursue, including in some cases getting married and raising a family, or spending the weekends getting drunk, or working in the family business, or going to church on Sundays, are such that the agent himself pursues them without desiring them or thinking them valuable: he is merely acting in accord with an inherited theory, doing the done thing. (9)

Even if someone could do the listed things *entirely* “without desiring them or thinking them valuable” *in themselves*, in any way at any time (which seems far too general to be plausible), then in order to have a motive he must desire and value “doing the done thing”.

Then there is an analogous account:

Similarly, many of the things an agent holds to be valuable, including things he holds to be valuable for himself, will be such that he has never questioned whether they are in fact valuable: he is merely taking on trust the truth of a theory handed down to him from his elders and teachers by the varied processes of cultural transmission.

Even if it is possible that someone has “never questioned” the valuable nature of some things that he “holds to be valuable”, in any way at any time (which seems implausible: introspection reveals us to doubt things quite readily and easily), then when he values them he will thereby have relevant desires, or wants, about them. Such questioning is not required by instrumental rationality.

However, the critique later incorrectly states that “Lester deems adherence to inherited theories rational because he claims they have withstood the test of rational assessment” (10), and he quotes *EfL* as a purported illustration of this assertion:

people must clearly perceive certain advantages in traditions, evaluate them as superior, if only in terms of the costs and benefits associated with those who keep them and the costs and benefits associated with those who break them. It is true that most people do not go in for radical criticism of all customs or habits they practise. They often give very little consideration to some of these, its being sufficient that they are content with them and see, on occasional reflection, no advantage to mending, at a cost and some risk, what does not seem broken (2012, 31-32).

This is then rejected by the critique for making two false assumptions: (1) “the possibility of a rational assessment independent of unexamined inherited assumptions”; and (2) “that such an assessment, in however minimal a form, has been completed”.

(1) As the quotation shows (thereby illustrating the value of quotation), *EfL* is making a point about traditional *practices* (“customs or habits they practise”); i.e., such things as marriage, celebrating Christmas, political elections, shaking hands, etc. These involve *intended actions* that can be given *instrumentally* “rational assessment”. The critique is extending the clearly cited sense to include all “inherited assumptions” and all “rational assessment”. But “inherited assumptions” of beliefs are not intended actions and “rational assessment” of beliefs is not instrumental (except insofar as thinking is a form of ‘action’). Therefore, it is invalid to criticise the instrumentally “rational assessment” of an intended action because it involves “inherited assumptions” that have not themselves been given (because they are not intended actions) instrumentally “rational assessment”. This is reminiscent of the invalidity of the criticism that the empirical-falsifiability criterion of being a science is self-refuting because it fails to be itself empirically falsifiable: it is invalid because the posited criterion is a philosophical theory and not itself a scientific theory.

All that said, it is not clear that we can even have *completely* “unexamined inherited assumptions”. For to use an “inherited assumption” is to assume that it is in some way correct (whether this means true, moral, prudent, etc.) rather than incorrect, and that seems to require some “assessment ... however minimal”: if only seeing that it does appear to us to be correct (hence we use it) rather than incorrect (in which case we would not use it).

(2) At least some “minimal” instrumentally “rational assessment” must have been “completed” in the sense that the agent deems it enough for action rather than further thought. If it had not, then the agent would be moving without having any idea of what he was trying to achieve and why; and that would not be an action. Of course, as *EfL* assumes critical rationalism, all assessments are held to be conjectural and provisional. In that sense they are never ‘justified’ and “completed”.

## “6. Self-Interest”

In *EfL* the interpretation of ‘self-interest’ as all “self-perceived interests” is simply explained and shown to be compatible with genuine altruism, including by a thought experiment. Every agent perceives that he takes an interest in all manner of things. Some of these are—ultimately, even if

involving other people, etc.—narrowly self-interested; relating to his own pleasure, happiness, status, health, etc. (although an agent can be mistaken about the consequences and what is really best for himself). Some of these are interests in the interests of other people or projects as ends in themselves; relating to family, friends, charities, science, etc. (although an agent can be mistaken about the consequences and what is really best for these). In order to be motivated, an agent must feel utility at the thought of promoting the interests of other people and projects. But these are altruistic to the extent that he would not wish or choose to lose these interests in other people and projects (by somehow forgetting them or ceasing to care about them) even if he knew that he would thereby gain more personal utility.

The critique goes back to abandoning quotation in favour of tortuous attempted paraphrase. Responses will be made as this proceeds.

I will state Lester's point as follows. In addition to his first-level interests, each person also has second-level interests, that is, interests in what first-level interests are fulfilled.

*EfL* does not mention the "level" analysis at all in this context.

When people act consciously, they are pursuing their second-level interests.

That seems to be acting self-consciously (or self-reflectively). First-level interests are still conscious.

Psychological egoism states that each person's second-level interests are concerned only with the fulfilment of his own first-level interests.

No, psychological egoism states that each person is necessarily concerned only with his self-referential interests (at whatever level); other people can only ever be means to his ends.

Psychological altruism states that some people's second-level interests are sometimes concerned with other people's first-level interests.

No, altruism is being concerned (at whatever level) with the interests of other people (at whatever level or none) for their own sakes.

Lester's thesis of self-interest says that each person pursues his own second-level interests.

No, people pursue those things they have a self-perceived interest in (at whatever level). Some interests will ultimately be about themselves (self-referential) even when involving other people. Some interests will be about other people for their own sakes, or as "ends in themselves" (which expression the critique finds "obscure" for some unexplained reason).

This is consistent with psychological altruism because some people may have second-level interests in other people's first-level interests.

No, it is consistent with altruism because people can have a perceived (any-level) interest in the interests of other people for their own sakes (at any of their levels—or none: because they are perceived not to know what is good for themselves).

Thus, second-level interests belong to the self but they may concern others.

As may first-level interests concern others (as ends in themselves or as a means).

It seems clear to me that this talk of 'interests' is far too amorphous and woolly to be useful if our aim is to obtain anything approaching a clear or precise understanding of agency or instrumental rationality.

As George Berkeley (1685-1753) observed in another context, "we have first raised a dust and then complain we cannot see". *EfL* does not give anything whatsoever like this concocted "amorphous and

woolly” and completely inaccurate “level” account of how people are always and only motivated by their self-perceived interests (in the self-perceived circumstances at the time) but are not thereby psychological egoists. Instead of quoting and dealing with *EfL*’s quite straightforward actual account, the critique’s earlier “classification” (four numbered distinctions in section four) is asserted “to be far more illuminating because it permits us to distinguish cases that Lester’s scheme lumps together” (11). The various inadequacies of the critique’s “classification” have already been dealt with in section four.

The critique then continues that “to use the same term for different cases slurs over differences that can be important. It achieves a simpler theory but, rather than advancing our understanding, it seems a barrier to it” (11). *EfL* is not replacing all different kinds of motivation with a single one. It is only explaining how these different kinds of motivation can still be self-perceived interests that can fit in the framework of instrumental rationality and utility-maximisation. Hence, this is “advancing our understanding” of how these things fit economic analysis and can thereby be related to want-satisfaction welfare and interpersonal liberty; which is the philosophical problem that *EfL* chapter two is trying to solve. As an analogy, suppose that the problem were to see how DNA-analysis explains all life-forms (on this planet, at least). It would be similarly confused to complain that “to use the same term for different cases slurs over differences that can be important. It achieves a simpler theory but, rather than advancing our understanding, it seems a barrier to it”. Of course, the critique just does reject the possibility of an instrumental-rationality analysis of all actions (analogous with rejecting a DNA-analysis of all life-forms), but none of the given criticisms appears to withstand scrutiny.

We are then given five possible ways of explaining why “Fred goes out of his way to cheer up a friend, despite the fact that it means missing a concert for which he has a ticket” (11). All this simply repeats the critique’s “classification”, which has been answered. But then the critique also posits the responses that are taken to be consonant with *EfL*. Replies will follow the quotations.

What does Lester say about these different cases? In each case he says that Fred acted in his self-interest.

No, *EfL* implies that Fred necessarily acted on a self-perceived interest, but not all interests *of* the self are interests *in* the self: for some interests are about other people (or other things) as ends in themselves.

Fred’s helping his friend was in his self-interest

It was one of his self-perceived interests, whatever the reason it was done.

because it was what he most desired (in (A)), or most valued (in (B)), or aimed to do (in (C) and (D)), or felt obliged to do (in (E)).

Except that, *a priori*, these must all be what in some sense he has the strongest overall desire (or want, or preference, or value) to do in the perceived circumstances at the time—or he would not attempt to do them.

What does it contribute to our understanding...

To “our understanding” of which *problem*? For *EfL* it is the problem of how an *a priori* core of instrumental rationality can always relate to agency (and thereby to economics, want-satisfaction welfare, and interpersonal liberty).

to be told that there is a flimsy (indeed, vacuous) level of description at which all these different cases can be described in the same way?

It solves the real and important philosophical problem. It is no more “vacuous” than DNA-analysis of all Earth life-forms, or mathematical-analysis in game-theory, or—more relevantly—‘aprioristic’

Austrian economics. Moreover, there is also the reply given to David Ramsay Steele's similar criticism that this is "all too promiscuously applicable" (*EfL*, 46):

On the contrary, action that does not fit the supposed end-means scheme is primarily begging for the, possibly difficult, task of suggesting a plausible, testable, enlightening new theory of the motive in operation. The alternative too often is idly, promiscuously, and unenlighteningly to assert some non-specific 'irrationality'. (*EfL*, 46)

But the critique answers, "Nothing that I can see". Because it ignores the explained philosophical problem and how the offered solution is supposed to work (just as the liberty critique does). Instead, we are given common-sense, colloquial classifications that are presented as essentially incommensurable.

The critique goes on to state that in "response to a somewhat different complaint from Amartya Sen (1977) that self-interest theories of action are vacuous, Lester asks three rhetorical questions" (12), which it quotes:

How can we choose to do what we do not in some sense prefer to do? Must not the chosen alternative be better for us in *some* sense? Otherwise, where is the personal motivation? (2012, 42)

The critique asserts that "these questions, intended as rhetorical, have more or less obvious answers contrary to those assumed by Lester" (12). Let us see whether the critique's "more or less obvious" 'rhetorical answers' withstand scrutiny. As usual, relevant quotations will be followed by replies at suitable junctures.

First, I can do what I do not prefer to do, in the sense that I do not desire to do it, or in the sense that I do not value doing it.

Which question is this supposed to be answering? Only the first question or all of them? Let us assume it is the first question. The first question asks, "How can we choose to do what we do not in *some* sense prefer to do?" The critique simply does not answer that question. It is completely irrelevant to that question to assert that there is *also* a "sense" in which "I can do what I do not prefer to do". *EfL* does not, and need not, deny that there are colloquial, common, and *other* senses in which we do not do what we prefer. It is making an *a priori* point that we must "in *some* sense prefer to do" what "we choose to do" (and hence instrumental rationality, etc., is possible). Similarly, of course, there is also "*some* sense" in which one must also "desire" and "value" doing what one does, and it is irrelevant that there are *other* senses in which one does not.

Of course, if one stipulates, as Lester seems to do (2012, 42), that to act is to reveal a preference, then there will always be that sense in which whatever I do is what I prefer to do; but that is trivial.

What one "stipulates" to be the case is thereby not a substantive part of any theory. It is held to be so by stipulative definition. *EfL* does not stipulate that "to act is to reveal a preference". It observes that there appears to be a "sense" in which this is *a priori* true. And what is *a priori* true is not thereby "trivial". In this case, we have an apparent *a priori* truth that allows for the instrumental rationality of all agency that economics can use (as can related theories of welfare and liberty).

Second, an agent may choose to do something that is less valuable for him than an available alternative because it is more valuable for someone else (or for some other reason or none).

Rather, "an agent may choose to do something that is less valuable for him [egoistically] than an available alternative because it is more valuable for someone else [whom he altruistically values more than his egoistic value in this case]". What about, "(or for some other reason or none)"? It just looks unintelligible to imply that an agent need not in any way value what he chooses to do, or may even have no reason whatsoever to do it. It appears to be *a priori* true that an agent *qua* agent must have a reason to act (or it would not be an action) and that he must most value that action (or he would have attempted the one he valued more).

Third, we often act intentionally without motivation, in that we neither desire nor value what we do.

Again, there are some senses in which “we neither desire nor value what we do”. Those senses would include not feeling an appetitive “desire” (or lust) and having no higher-order (second-level) “value” (including moral principles). But those senses are irrelevant to the *a priori* point that intentional acts must also in “*some* sense” be desired (or wanted) and valued (or thought good).

Further, when we do desire or value what we do, the desire or valuation does not make us do it: our intentional actions are undetermined at every time before they are begun.

The irrelevance of the determinism-indeterminism debate has already been explained. This is about what is *a priori* the case, not what is causal. However, it seems worth noting here that adherence to a “free will” that rejects compatibilism, determinism, and random-indeterminism without offering any account of how that could operate, is another example of the incoherent background assumptions of the critique.

Lester’s treatments of the objections of C. D. Broad and Tibor Machan (2012, 43-46) seem similarly question-begging.

What does the critique mean by “question-begging”? Presumably, in this context, this is a charge of *petitio principii*. But all valid arguments are *petitios* (the conclusions have to be implicit in the premises), and so this is not a valid criticism. There is sufficient detail in the replies to Broad and Machan that they could have been properly criticised instead of being dismissed in this cursory and invalid way.

## “7. Maximisation”

In *EfL* agents are explained and defended as utility-maximisers in the *a priori* sense that they attempt to take the action that maximises their utility (or want/preference-satisfaction) at that time in the circumstances as they conjecturally perceive them to be. This core sense thus differs from mainstream economics, which interprets utility-maximisation as a successful long-run achievement concerning all an agent’s preferences (fully realising that this assumption of perfect calculation over time can only be a useful, broad approximation to the truth).

The critique quotes *EfL* on how “we seek to maximise our want-satisfaction”:

‘as we compare possible choices we cannot help but take the option that in some way feels to be the most want-satisfying, or least want-dissatisfying, at the time.’ ... (2012, 50-51).

Two criticisms are then given:

First, most of our decisions are habitual or conventional rather than reasoned ... , so what is chosen is the usual rather than the best or the most want-satisfying.

A first thing to note is that there is simply no inconsistency between choosing what is “habitual or conventional” or “usual” and what is “the most want-satisfying” thing at that time; and therefore the “so” is an invalid inference. In fact, choosing in this way is often “the most want-satisfying” thing to do. And where it is not, we choose in the way that is. As *EfL* observes and discusses, this is “introspectively knowable” (50). Otherwise, we would have to say that habits, etc., sometimes bypass our wants and cause our bodies to move like puppets. No very great self-consciousness is required. So, it might superficially appear—especially after the largely-forgotten event—that one ‘automatically’ or ‘unconsciously’ made a cup of tea first thing in the morning as usual. But if one were to consider, or be asked about, one’s behaviour at the time, one would immediately see that this habit is being followed because it is more want-satisfying, or preferred, than any alternative action that comes to mind (including simply not doing it).

The critique's use of "reasoned" is apparently intended to imply significant deliberation. But all thought—however superficial, panoramic, unselfconscious, and non-linguistic—is 'reasoning' in a more general sense of that term. Thus, we cannot engage in what is "habitual or conventional" without having "reasoned" about what we are doing in that sense. As in so many similar cases, with "reasoned" the critique does not capture some narrow, essential, correct usage.

Second, even in cases where we reason about options, we usually cannot identify all the options, all their consequences or all the relevant evaluative principles ... , so we often cannot identify the best, or the most want-satisfying.

But the instrumental rationality thesis being defended only holds that we choose the "most want-satisfying" from among the options that we perceive at the time. So, this criticism is the usual, unwitting, attack on a straw man.

The critique cites Simon 1997's "bounded rationality" and the idea that we "'satisfice,' or look for a course of action that is good enough, rather than seeking to maximise, or look for the option which is best". The critique mischaracterises *EfL*'s response, writing that "apparent satisficing is really maximising because, in such cases, at some point we guess that the disutility of search costs is likely to outweigh any other utility that we will achieve"; and the critique says that "it seems to be a wholly inadequate one" (13). First it is asserted that we "saw this in connection with habitual and conventional actions". But that has already been answered. The following additional reasons are then given (and will be replied to here, as and where seems appropriate):

First, suppose that Lester were right that the decision-maker is able to say that further search will be more costly than it is worth.

*EfL* does not say that it "will be". In terms of the *a priori* instrumental rationality being considered, it is not a matter of whether the decision-maker is able correctly to predict the consequences of his decision. He makes the judgement, however cursorily, *now* that as he considers what he conjectures to be further search costs and the likelihood of a better outcome, his *current* most want-satisfying (or utility-maximising) option is to stop searching and act.

This does not alter the fact that the decision-maker would still not have enough information about the options before him to be able to say which is better than the others.

As we always operate within a framework of conjectures, and as the potential information is infinite, and as we never know what relevant possibilities we might have overlooked, then it is in principle impossible in any objective and supported sense to know that we have "enough information about the options". It is, as we have seen, also irrelevant.

Second, when Lester says that the decision-maker guesses that further search would not be worthwhile, he might mean that the decision-maker just makes an arbitrary, unreasoned decision to search no farther.

No. A guess, or conjecture, aims at a certain outcome and is shaped by critical considerations, insofar as these occur to us. Thus, it is never "arbitrary" or "unreasoned".

Third, if Lester means ... that he calculates whether further search is worthwhile, then we get a vicious circle.

And then the critique goes, yet again, into fanciful and elaborate "first-level options" and "second-level options" that he asserts, wholly incorrectly, to be "Lester's view". It is not held, or logically implied, in *EfL* that the decision-maker inevitably "calculates" in any complicated way, although he may if he prefers that option. At some point, he simply finds that the thought of acting now gives him more utility now than engaging in further search costs. Of course, in terms of his overall long-term utility he might be making a mistake. But the *a priori* instrumental rationality that is being defended is not about efficiently maximising long-term utility.

## “8. Morals”

In *EfL* morals are given a formal interpretation that explains how moral beliefs or sentiments fit instrumental rationality, rather than being outside it or conflicting with it. In short, the categorical nature of moral obligations means that, when they are genuinely held, they necessarily override all non-moral goals in terms of the utility from keeping them and the disutility from breaking them.

A first important thing to note is that the interpretation of the categorical nature of moral obligations defended here is not necessary for making sense of *a priori* instrumental rationality. It would be sufficient that different moral obligations were ascribed variable amounts of (dis)utility, and that these compete with the (dis)utility of the other choices that an agent has. That said, if we assume that people can simply choose to act immorally, then that would tend to undermine the extent that it is thought worth attempting to reason with them about what is moral and immoral (and perhaps thereby even discovering that they are right and that we are mistaken). Dogmatism and coercion become the default positions, which will tend to militate against liberty and welfare (thereby being both undesirable and not fitting the “compatibility thesis” of *EfL*).

The critique attempts to paraphrase *EfL* on how morals fit into instrumental rationality, and then selectively quotes *EfL*:

fully to hold a moral obligation sentimentally, not to feel it uncertainly or as a slight pricking of the conscience, is always to act on it in appropriate circumstances...It is possible to defend moral theories intellectually without really feeling them. Without seeing this, one can fail to realize that one’s ‘official’ or ‘theoretical’ moral position is a sort of public recommendation that one might not personally feel, value, or desire...we cannot knowingly do what we feel, at that moment, is immoral (2012, 51-52).

As usual, the critique responds with more common sense than philosophy (responses will follow the quotations):

Lester’s position here seems to fly in the face of human experience.

This “human experience” seems on a par with the critique’s earlier “descriptions of ordinary facts of life”. It is the problematic, philosophically unexamined, way that things are usually seen.

I seriously doubt that that there has ever been, or will be, any person who has not on many occasions acted in a way that he is at the time of acting convinced is wrong.

This is emphatically to reassert the common-sense position instead of criticising the philosophical arguments that *EfL* presents.

One type of case involves weakness of will,

Which myth has been dealt with already (section three).

but in many cases there is no such weakness because the agent is resolute in pursuing the course he is convinced is wrong and perhaps steels himself to do it.

As this fails to directly address the arguments in *EfL*, the most useful response is probably to repeat some of them. It is generally accepted that moral theories are about categorical—or absolute—limits on acceptable behaviour, irrespective of any other conflicting aims an agent might have. This categorical feature is, *ipso facto*, never overridable. If there are perceived extenuating reasons to do something that is generally described as “immoral” (such as a clash with a more important moral principle, or some great evil that would otherwise occur), then it is perceived as not actually immoral in those circumstances after all. If behaviour that is immoral (“wrong”) is behaviour that is absolutely unacceptable/impermissible/undesirable, then it is *a priori* necessary that someone cannot sincerely hold, at time T, that X is immoral (i.e., feel it to be absolutely unacceptable) and also, at time T, try to

do X (thereby implying it is felt to be acceptable). The critique does not criticise this philosophical line of reasoning, it merely confidently asserts the common-sense position that is inconsistent with it.

The varied situations and ways in which this happens has been explored extensively in literature, and also in film,

That the common-sense position is accepted in popular culture is not a philosophical argument against *EfL*'s position.

and its effects have been analysed in some psychological studies of cognitive dissonance

Something or other may be “analysed” in those studies. But without a coherent explanation of chosen immorality, that is like saying that the effects of seeing square-circles “have been analysed in some psychological studies of cognitive dissonance”.

Typically, the person who knowingly does wrong

There is still no explanation of how this is coherent.

tries to dissipate the discomfort he feels by seeking a way to justify himself: ‘It is not so bad,’ ‘the circumstances are exceptional,’ ‘I have a good excuse,’ ‘other people would do the same in my position,’ and so on.

But some of those reasons will sometimes be sufficient to persuade the agent that, despite the general moral rule, his circumstances do constitute a genuine exception. And some such reasons must be believed if he is to decide that what at first might appear to be completely unacceptable is actually acceptable in his circumstances. Alternatively, as the quotation of *EfL* says, it is possible that he doesn't genuinely believe his “official” or “theoretical” moral position. Due to confusion, he might even genuinely believe that he believes it, but that is not the same thing as genuinely believing it. Actions speak louder than words here: when people act, it will tend to be on the basis of what they really believe.

He tries to delude himself into thinking that he has not really done wrong at all.

Note first this switch to after the event: “he has not really done wrong”. Of course, someone can change his mind about the moral status of his action even moments afterwards. That said, it seems—by introspection—impossible that we can intentionally delude ourselves in the sense of making ourselves believe what we do not believe (try to believe that you are Napoleon, or can fly by flapping your arms). It is possible to try, but it will be in vain. However, it is quite plausible that, around the time of the action, the agent is seeking sound reasons to believe that he is not about to do wrong. And at the moment of his action that is what he must believe—however influenced by emotion his reasoning might be—in order to overcome the moral view that his behaviour is completely unacceptable. He will always feel that whatever he actually does is fully acceptable (in fact, the best thing to do in the perceived circumstances) and thereby not immoral (categorically unacceptable) at all.

This, indeed, may be the truth behind Lester's claim: not that no one knowingly does wrong, but that most people who knowingly do wrong succeed in deluding themselves that they have not done so.

All this has now been explained, but it might help to recapitulate. It is, on analysis, *a priori* incoherent to assert that someone “knowingly does wrong”: by understanding that immoral behaviour is categorically unacceptable and what that logically entails if a moral view is genuinely believed. Moreover, no one can choose to delude himself. And it is, in any case, irrelevant that we might assess the morality differently *after* the act (“that they have not done so”).

The critique then asserts, without a page reference, that “Lester refers to Hare for further argument for his [Lester's] position” (15). However, *EfL* rejects Hare 1952 on this matter, despite

agreeing with parts of its moral theory (*EfL*, 53, 209). The critique then digresses into a paraphrase of Hare 1952's moral theory and adds criticisms of it. It is not relevant to defend Hare 1952 itself here, but—in the interest of replying comprehensively to the critique in *EfL*'s terms—some tentative attempts at responses will follow quotations of the criticisms.

the objection I want to raise here is that we can discuss dispassionately whether a thing is good or right, without making any prescriptions;

Of course, if we fail to reach any conclusions. Or, perhaps, if we are interested in which things are, in themselves, “good or right” rather than prescribing what ought to be done about them.

indeed, we can even prescribe what we think is not good or not right.

To *sincerely* “prescribe” a moral view seems to imply that the ends are, overall, perceived by the prescriber to be “good” and that actively promoting them is “right”.

Hare's response to this objection is to say that, in such cases, we are not making genuine value-judgements, but are rather making an ‘inverted-commas use’ of the terms ‘good’ or ‘right,’ that is, we are using the terms to refer to the non-value properties which incline some people to call things ‘good’ or ‘right’ (1952, 163-70).

If this means that we can insincerely prescribe things according to how other people use ‘good’ or ‘right’, then that makes sense. We must think that such insincerity is not immoral.

However, Hare is here merely re-describing the examples to fit his theory.

To show how putative counterexamples can be interpreted to fit a theory that is being defended is not thereby to engage in “re-describing the examples” in any invalid way.

Perhaps this can be seen most clearly if we consider a character like the devil, who may prescribe what he thinks to be bad or wrong.

Rather, the devil will prescribe some things that are “bad or wrong” for humans because he perceives those things as good and right in his terms. He has different moral values.

Hare has to treat this as an inverted-commas use of ‘bad’ or ‘wrong’ (1952, 175). But that is mistaken, since what the devil intends to prescribe is not whatever some people consider to be bad or wrong; rather, being the devil, he intends to prescribe what is really, objectively, bad or wrong.

Rather, the devil “intends to prescribe what is really, objectively, bad or wrong” for people because he perceives it as good and right for him to do this.

If he says ‘what is good is bad,’ he is using one term descriptively, to talk of what is good (or bad) and the other term to deprecate (or commend);

If the devil means that what is good for people is perceived as bad by him (or what is perceived as good by him is bad for people), then that seems to make sense. Similarly, what is good for the greenfly is perceived as bad by the gardener.

and similarly if he says ‘the right thing to do is the wrong thing to do.’

If the devil means that what is “the right thing to do” for people is perceived as “the wrong thing to do” by him (or, again, *vice versa*), then that also seems to make sense. Similarly, avoiding the slug traps is the right thing to do for the slugs but the wrong thing to do as perceived by the gardener.

We can understand this easily enough, because there is a bit of the devil in each of us, though some people seem to have more of it than others.

Rather, we can understand this easily enough because we can see the different perspectives of different agents.

Thus, while it is true that terms such as ‘good’ and ‘right’ may be used by speakers to prescribe, we should acknowledge that this is a secondary use, their primary use being to ascribe a value property.

Possibly true, but not relevant to the real dispute.

Hare’s view is mistaken because it omits mention of the properties of goodness and rightness.

As they are described, the “properties of goodness and rightness” are relative to the different values, ends, and perceptions of the different agents. (This is not to deny that there could be a correct theory of “goodness and rightness” that transcends individual perceptions of these things.)

Back to the critique of *EfL*:

As so often, Lester attempts to bolster his case by asking a rhetorical question: ‘can you recall doing anything that clearly felt immoral at the time that you did it?’

As so often, the critique attempts to bolster its case by ignoring the details of the surrounding arguments in *EfL*, seizing on some illustrative question that is asked as a serious test, and then calling it “rhetorical” and giving a common-sense rhetorical answer to it.

His assumption, of course, is that the answer is ‘no.’

In fact, *EfL* explains why the answer should be ‘no’, but clearly assumes that the question is always likely to be answered ‘yes’ if faced with common-sense assumptions that ignore the explanation.

However, my answer is ‘yes.’

Quite.

I can recall many examples of my doing things that I clearly thought were immoral.

That means “doing things” that were sincerely believed at the exact time of the chosen actions to be categorically impermissible/unacceptable/undesirable chosen actions. How can that make sense? (That question is not rhetorical.)

Anyone else could do the same, provided he has not deluded himself with the self-justifying chicaneries of cognitive dissonance.

Anyone else might be equally confused by common-sense assumptions, but he cannot choose to delude himself. And being “deluded” *after* the event would be irrelevant to the thesis, in any case.

On Lester’s mechanistic theory of motivation, the supposed fact of the impossibility of doing what one knows to be wrong is explained by positing that moral desires are always stronger than non-moral desires (2012, 52). He says:

what is felt to be immoral is what we feel no one should ever do in the circumstances; it is a categorical sentiment... (17)

And here the critique’s ellipses cut out and ignore the explanation of the categorical nature of morals:

about a type of behavior in some group. One cannot at the same time (at least, not without confusion) do what one feels no one should do. Moral values must be obeyed because if disobeyed they are, *ipso facto*, not held categorically. Most moral philosophers seem to agree that morals are at least categorical whatever else they are. (*EfL*, 52)

Instead, the critique jumps to,

One source of confusion here is where our general moral feelings (such as feeling that lying is usually immoral) differ from our specific moral feelings (such as feeling that some particular lie is moral). This seems to occur because such general morals are usually held *ceteris paribus* (2012, 52-53). (17)

And the critique's reply (with responses made here) immediately follows.

Is it possible to have a feeling with so reticulated a propositional content

The feeling does not have a propositional content but is a response to an idea. It is a feeling of complete rejection at the thought of a certain type of behaviour. We then typically call that behaviour 'wrong' or 'immoral'. Similarly, a feeling of belief might be the response to the thought of some descriptive proposition. We then typically call that proposition 'true'. Or a feeling of delight might be a response to a passage of music. We then typically call that music 'beautiful'.

as no one should ever do an action of type A in these circumstances (where an appropriate action-description is substituted for 'A')

It seems clear enough that someone can have a feeling of complete rejection (categorical impermissibility) for an action of type A in circumstances X (e.g., A: punching someone, X: when we do not like the look of him).

or lying is usually immoral

The feeling of complete rejection is a response to the thought of typical examples of lying.

or, even more implausibly, *ceteris paribus*, no one should ever do an action of type A in these circumstances (where an appropriate action-description is substituted for 'A')?

The thought itself may be as complicated and qualified as someone is capable of entertaining.

The supposition seems ridiculous.

*EfL* does not assert or imply the "supposition" that the feeling of complete rejection itself includes the idea to which it is a response.

Of course, one might say, 'I feel that no one should ever do that sort of action in these circumstances.' But there one is not using 'feel' literally, to talk about a desire or sentiment, but to indicate one's uncertainty about the thought one is expressing.

This, typical, lapse into the common-sense usage of colloquial language is, typically, entirely beside the point. If there are moral feelings about certain types of action—and it seems that there are—then one of those feelings can quite intelligibly be expressed by saying, "I feel that no one should ever do that sort of action in these circumstances". It is irrelevant that those words could also be given the interpretation that the critique does.

It is, indeed, possible to have an attitude toward a complex thought, for example, one might admire it or feel happy at the contemplation of it,

Or have a feeling of absolute rejection towards the type of action it describes.

but the feeling itself does not have the structured complexity of the thought.

And *EfL* does not say or imply that it does.

Lester seems driven to supposing that it does, not simply by his mechanistic theory of motivation,

There has to be felt motivation or we would not have anything to make us act. That does not make that motivation “mechanistic”. The critique appears to interpret all felt motivation as thereby mechanistic and so opts instead for an incoherent and mysterious motivation that is somehow supposed to be a pure “act of will”.

but by that in conjunction with his attachment to construing values as felt desires.

To value something must be to desire it in some respect: if not for oneself, then that it exist, etc. All real values must be felt. It is only a metaphor to speak of “values” that are not felt. For instance, “liberty is a value” cannot literally make sense. What social, or interpersonal, liberty is literally—on *EfL*’s abstract interpretation, at least—is the absence of interpersonal initiated constraints on want-satisfaction (see Lester 2012, 2014, 2016). One can certainly value such liberty, but liberty in itself does not thereby become a literal value (similarly, one can believe theory X, but theory X in itself does not thereby become a literal belief). The idea that one can genuinely hold any kind of value without having any desire as regards it is not intelligible.

## “9. Conclusion”

The critique’s “conclusion” summarises its criticisms of what it construes to be *EfL*’s arguments. Rather than here summarise all of the responses already made, a more general conclusion might be more useful.

If we were to take the critique completely seriously, then this would appear to give rise to one obvious and severe problem. Economics is about the efficient allocation of scarce resources when there are alternative uses and ranked goals.<sup>8</sup> But, as a social science, economics is also based on some interpretation of every person being an instrumentally rational ‘self-interested utility-maximiser’ (*homo economicus*). Abandoning this assumption as being realistic in any serious sense would appear to make economics as a social science impossible or almost uselessly abstract (it would be somewhat like the joke about the physicist’s farming advice that starts by assuming spherical cows in a vacuum).

Otherwise, this critique and the earlier liberty critique comprise the most sustained criticisms so far of the overall theory of eleutheric-conjectural libertarianism that *EfL* (etc.) propounds (although the current critique is less central and there is a great deal more in *EfL* than both critiques even touch on). If only for this reason the two critiques have some merit; and that they have, thereby, prompted further explication of the relevant parts of the theory. However, as both this reply and the previous one show, they fail even to demonstrate a grasp of either the relevant philosophical problems or the offered philosophical solutions. Unless and until this philosophical theory of libertarianism is given competent critical consideration, libertarianism will remain at the level of fundamental philosophical confusion that has barely made any progress since Locke’s *Second Treatise*.<sup>9</sup>

(2016; revised February 2022.)

---

<sup>8</sup> Thus interpreted, it can be applied to the evolved structure and behaviour of non-human animals and even plants: greater economic efficiency having a greater survival value.

<sup>9</sup> This essay has benefitted from comments provided by Mark Brady and David McDonagh.

## References

- Frederick, Danny. 2013. "A Critique of Lester's Account of Liberty." *Libertarian Papers* 5, 1: 45-66. Online at: [libertarianpapers.org](http://libertarianpapers.org).
- . 2015. "Adversus Homo Economicus: Critique of Lester's Account of Instrumental Rationality." [https://www.academia.edu/2037838/Adversus\\_Homo\\_Economicus\\_Critique\\_of\\_Lester\\_s\\_Account\\_of\\_Instrumental\\_Rationality](https://www.academia.edu/2037838/Adversus_Homo_Economicus_Critique_of_Lester_s_Account_of_Instrumental_Rationality)
- Hare, Richard Mervyn. [1952] 1986. *The Language of Morals*. Oxford: Oxford University Press.
- Lester, J. C. [2011] 2016. *Arguments for Liberty: a Libertarian Miscellany*. 2nd ed. Buckingham: The University of Buckingham Press.
- . [2000] 2012. *Escape from Leviathan: Libertarianism without Justificationism*. Buckingham: The University of Buckingham Press.
- . 2014. *Explaining Libertarianism: Some Philosophical Arguments*. Buckingham: The University of Buckingham Press.
- . 2017. *Two Dialogues: Introductions to Philosophy and Libertarianism*. Buckingham: The University of Buckingham Press.
- Percival, Ray Scott. 2012. *The Myth of the Closed Mind: Understanding Why and How People Are Rational*. Chicago and La Salle, Illinois: Open Court.
- Simon, Herbert. 1997. *Administrative Behavior*, 4th ed. New York: The Free Press.
- Watson, Gary. 1975. 'Free Agency.' *Journal of Philosophy* 72 (8): 205-220.