

## CRITIQUE OF LESTER'S ACCOUNT OF INSTRUMENTAL RATIONALITY

Danny Frederick\*

*Draft 25 June 2013*

**Abstract** In Chapter 2 of *Escape from Leviathan*, Jan Lester defends two hypotheses: that instrumental rationality requires agents to maximise the satisfaction of their wants and that all agents actually meet this requirement. In addition, he argues that all agents are self-interested (though not necessarily egoistic) and he offers an account of categorical moral desires which entails that no agent ever does what he genuinely feels to be morally wrong. I show that Lester's two hypotheses are false because they cannot accommodate weakness of will; because they are inconsistent with agency, which requires free will; because ends, obligations and values cannot be reduced to desires; and because maximisation is often not possible. Further, Lester's claim that agents are self-interested is vacuous, his attempted reduction of moral behaviour to want-satisfaction fails, and his contention, that agents always do what they genuinely think to be morally required, seems untenable. A defence of freedom that depends on *homo economicus* is far from promising.

**Keywords** agency; cognitive dissonance; desires; ends; free will; Hare; *homo economicus*; instrumental rationality; Lester; Leviathan; libertarianism; maximisation; morality; Popper; self-interest; values; Watson; weakness of will; wrongdoing.

## 1. Introduction

In Chapter 2 of *Escape from Leviathan: Libertarianism without Justificationism* (Lester 2012), the libertarian and Popperian philosopher, Jan C. Lester, offers a defence of the '*homo economicus*' conception of people as self-interested, utility-maximising, rational agents. Lester sets out his connected conceptions of agency and of instrumental rationality and he seeks to defend them against a range of criticisms. I will show that his defences often fall down, and I will offer additional criticisms that he does not consider. I show that the mechanical, *homo economicus* conception of rationality and of agency is thoroughly mistaken. I offer instead a conception of people as fallible, largely ignorant, often irrational and sometimes immoral, who inherit theories about facts, values, morals and themselves which they strive to improve in a piecemeal fashion by conjecture and experiment.

In section 2, I explain instrumental rationality and then Lester's thesis of instrumental rationality, which conjoins a hypothesis about rationality with a hypothesis about agency. In section 3, I argue that the phenomenon of weakness of will refutes Lester's hypothesis about agency and that Lester's attempt to explain away the phenomenon is unsuccessful. In section 4, I show that Lester fails to make sense of the conflict between values and desires, that his want-satisfaction account of agency should be supplanted by an account which recognises distinct categories of desires, values and ends, and that this scuppers his account of instrumental rationality. In section 5, I argue that Lester's passive conception of agency is not really a conception of agency at all, because it does not allow for metaphysical free will; and this entails the rejection of Lester's hypothesis about rationality as well as his hypothesis about agency. In section 6, I dismiss Lester's contention that all

---

\* Webpage: <http://independent.academia.edu/DannyFrederick>, Email: [dannyfrederick77@gmail.com](mailto:dannyfrederick77@gmail.com)

actions are self-interested (though not necessarily egoistic) as vacuous. In section 7, I argue that Lester's assumption that agents always seek to maximise the achievement of their ends is false and that this refutes both his hypothesis about rationality and his hypothesis about agency. In section 8, I show the falsity of Lester's contention that we cannot act against our moral convictions. In section 9, I conclude.

## 2. Instrumental Rationality

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, and it is instrumentally irrational if and only if it is an unsuitable means to the achievement of the agent's ends (see Popper 1994b, 178-79, 181, 183n19; and my 2013). As it stands, this statement may be interpreted objectively, subjectively, or part one way, part another. On the objective interpretation, instrumental rationality requires that the agent's actions are objectively suitable to his objective ends. An agent's objective ends are those that are assigned to him ontologically, by his nature, or by Nature, or perhaps by God. It may be disputed whether an agent, or at least a human agent, has any objective ends; but if he has, then it is possible that he is quite mistaken about what they are, and he may seek to achieve ends which are not only different to his objective ends but which even undermine them. An agent's subjective ends are those which he acknowledges as ends. An action is objectively a suitable means to a given end (whether subjective or objective) if and only if it is in fact a suitable means to achieving that end, whether or not the agent thinks it is. An action is subjectively a suitable means to achieving a given end if and only if the agent thinks the action is a suitable means to achieving that end. We can therefore interpret the minimal conception of instrumental rationality in three ways:

- (o) an action is (objectively) instrumentally rational if and only if it is a suitable means to the agent's objective ends;
- (h) an action is (hybrid) instrumentally rational if and only if it is a suitable means to the agent's subjective ends;
- (s) an action is (subjectively) instrumentally rational if and only if the agent thinks it is a suitable means to his subjective ends.

I ignore, as uninteresting, the fourth possible case, in which an action is thought by an agent to be a suitable means to an end which happens to be an objective end of his but which he does not acknowledge as an end.

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). He does not defend the egoistic version of this thesis: he concedes that people may often want to promote the interests of others. He does not interpret the thesis hedonistically or eudaimonistically: he accepts that people often want things other than pleasure or happiness. He also admits that people can make mistakes about their own long-term interests. In making this point it seems that Lester uses 'interests' to talk about an agent's objective ends, and uses 'wants' to talk about an agent's subjective ends. However, generally, when Lester uses 'interests' he is talking about wants or, as he sometimes puts it, 'perceived interests,' or 'self-perceived interests.' He regards it as *a priori* true that 'we are always motivated to satisfy our greatest desire, or want, or appetite, or perceived interest'

(2012, 16). He seems to use 'desire,' 'want,' 'appetite' and 'perceived interest' interchangeably.

Lester's *a priori* thesis of instrumental rationality is a conjunction of two hypotheses. The first is a hypothesis about rationality, namely, that instrumental rationality requires an agent to maximise the satisfaction of his desires over time, which entails that it is irrational not to maximise the satisfaction of one's own desires over time. This sounds like a hybrid hypothesis of instrumental rationality because it concerns subjective ends but it seems to be about actions which are objectively means to their maximisation. However, from Lester's statement that 'Decisions are always made on the *perceived balance* of costs and benefits' (2012, 25), it seems clear that he intends his hypothesis to be about subjective instrumental rationality, and thus about actions which the agent thinks are means to his ends (see also 2012, 15). We can state Lester's rationality hypothesis more clearly as follows:

(LR) an action is (subjectively) instrumentally rational if and only if the agent thinks it will maximise the satisfaction of his wants over time.

The content of (LR) consists of the conjunction of (s) with the following two postulates:

- (a) an agent's subjective ends are always the ends of his desires;
- (b) a means to an end is suitable if and only if it maximises the achievement of that end.

I will show that both of these postulates are false and that, consequently, (s) is superior to (LR).

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

(LA) all actions are instrumentally rational.

This is a bold hypothesis because it rules out irrational action. Given (LR), it implies that an agent always acts in accordance with his strongest desire. A qualification needs to be added. For example, an agent will not act to satisfy his strongest desire if he is chained to a wall and unable to so act. We might say that the action in question must be possible for him. But that is not quite right, because he still might not act if he is not chained to the wall but mistakenly thinks that he is. So 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. Some further tweaks may be needed to accommodate other *recherché* possibilities, but I will ignore all such qualifications in what follows. I will show that (LA) is false.

Lester defends his *a priori* thesis against a range of objections. I will argue that his defences do not succeed and I will offer some additional objections. The following discussion employs 'instrumental rationality' in the subjective sense throughout.

### 3. Weakness of Will

Lester's hypothesis about agency, (LA), implies that to act is to do what one most desires, most wants, or thinks it best to do. Yet, cases of weakness of will – for instance, someone wants to stop eating chocolate but succumbs to the temptation to

eat another bar – involve someone wittingly doing something other than what, under the circumstances, he most desires, most wants, or thinks it best to do. Lester therefore declares that there are no real cases of weakness of will (2012, 24). This is a problem because cases of weakness of will seem familiar to us all from everyday life and have been described convincingly in many great works of literature (for example, Tolstoy 1911). Lester tries to dissolve the problem by explaining away apparent cases of weakness of will. For this purpose he invokes Harry Frankfurt's (1971) distinction between first-level desires for actions and second-level desires to have, or not have, first-level desires.

Lester's exposition (2012, 24-26) is not entirely clear, but I think his suggestion is that what is really going on in a case of weakness of will is that:

- (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;
- (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;
- (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.

For example, a 'smoker might like not to desire smoking so much, but given that he does desire it so much he regards himself as better off by smoking than by not smoking. He might want to cultivate a stronger desire to stop because, for instance, he wants to live a little longer' (2012, 24-25). Similarly, 'A woman strongly desires chocolate and feels that life without it is too miserable to forgo it.' However, 'She knows that chocolate makes her fat and feels that being fat is worse than losing her desire for chocolate' (2012, 26).

However, this alternative account of apparent cases of weakness of will is not convincing. 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. 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. 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. Lester says that when people make such assertions they are being insincere, just giving a sop to their critics (2012, 25). That might be true in some cases, but it seems to be false in others. 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. For example, someone who wants to give up smoking may plead with his friends and family to prevent him from getting hold of any cigarettes.

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, 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. In consequence, the connection Lester proposes between (ii) and (iii) does not hold. Even if, as (ii) requires, her first-level desire not to be fat is stronger than her second-level desire to retain the desire to eat chocolate, it need not be the case, as (iii) requires, that her strongest second-level desire is the desire not to have the desire to eat chocolate.

The phenomenon of weakness of will refutes (LA).

#### 4. Desires and Values

Gary Watson, invoking the Platonic distinction between reason and desire, says that desires and values constitute two different sources of motivation, the former being non-rational, the latter rational. There is therefore a difference between desiring something and thinking it to be of value (or judging it to be good): some things we desire, we do not value, or the strength of the desire may be out of proportion to the thing's value. Only values are, in themselves, reasons for action (1975, 208-213). For example, a woman who has a sudden urge to drown her bawling child in the bath does not value her child's being drowned; and a man who thinks his sexual inclinations are the work of the devil...bespeak[ing] his corrupt nature...does not acknowledge even a *prima facie* reason for sexual activity; that he is sexually inclined toward certain activities is not even a consideration (Watson 1975, 210).

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. 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.

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. 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. 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. 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.

Lester seems reluctant to accept that there are objective values, 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, he offers a different account, one which invokes, again, the distinction between levels of desire. On this account, the woman has a desire to drown her child and a second-level desire not to have that first-level desire; and the man has a desire for sexual activity and a second-level desire not to have that first-level desire. But this account fails for the same reason that Lester's account of weakness of will failed. The man, for instance, may strongly desire that he retain his desire for sexual activity, for he may see it as a test of his virtue or strength of character, so that abstaining from sexual activity despite a desire for it has much greater moral worth than abstaining from sexual activity merely because one has no desire for it.

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.' Whereas Lester seems to see all desires as appetites (2012, 16), Watson distinguishes appetitive or passionate desires from those which are the products of culture or habituation. 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 (1975, 214-15). Both Lester and Watson think that action results from motivation, and that motivation depends on desires and values. It seems to me that there are many errors in these views.

I suggest that we distinguish:

- (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.

Theorists have tended to conflate these, but they are distinct. A felt desire is one that an agent feels, such as a hungering, a craving, a yearning, a longing, a lust or an urge. This seems to be the primary sense of 'desire.' An agent values something if he thinks it is valuable for him or valuable in general. This seems to be the primary sense of the verb 'values.' In these senses, 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; or I may go to work despite feeling no desire to do so, because I think going to work is valuable for me. 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; 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. However, I may also both desire and value something, as when I eat because I have a healthy appetite. 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. 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 person may have no felt desire for classical music, despite valuing it. Although 'desire' applies primarily to (1) it is also applied metaphorically to (2) and (3); and although 'values' applies primarily to (2) and (3), the term has an analogical sense in which it applies to items in (1) too.

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. 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. 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. Yet 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 that I might fail to pull off, for example, by missing my mouth and spilling the juice all over my shirt; 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 (see my 2010a for discussion and other examples, and section 5, below, for some more substantial examples). So, (4) are distinct from (1), (2) and (3). Nevertheless, in practice, the term 'desire' or 'want' is extended to such cases. For example, 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. The agent's desire to perform the action *in this sense* simply follows from the fact that the action is intentional (Nagel 1970, 29-30). This is an even more attenuated sense of 'desire' than that in which it is applied to (2) and (3).

The fact that agents may have ends which are not ends of their (felt) desires refutes postulate (a); it also refutes Lester's rationality hypothesis, (LR), since postulate (a) is part of the content of (LR). The refutation goes through even if we weaken the content of (a) so that 'desire' covers values as well. Postulate (a) would be saved from refutation if 'desire' were weakened further, to cover also aims; but it would then become vacuous, saying only that an agent's subjective ends are always that agent's aims. It also seems clear that Lester endorses the strongest of these three versions of (a).

## 5. Free Will

Lester, like Watson, employs a passive conception of agency. For Lester, each agent has a set of wants or desires, and these, in conjunction with the agent's hypotheses about the facts of the world, bring about the agent's actions. The agent always performs the action that he thinks will most satisfy his wants. For Watson, an agent has desires and values. So long as the agent is rational, his values, in conjunction with his beliefs about the non-value facts of the world, bring about his actions. When he is irrational, desires of the agent which are at variance with his values, in conjunction with his beliefs about the non-value facts of the world, bring about his actions. On these views, human action is explained in terms of 'motivation' which impels an agent to act.

It seems clear to me, however, that this view of agency is not a conception of agency at all. For, on such a view, 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. 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. Similarly, Lester says: 'Free will seems to require only choice, and human beings always have to choose their movements except for certain autonomic ones' (2012, 21-22). The problem is that, if actions are always dictated by the strongest desire, the agent has no choice.

Lester says he uses 'free will' in the sense of acting without being forced by another (2012, 20). This interpersonal sense of the term is central in some discussions in moral, legal and political philosophy, but there is another sense of the term which is also relevant to those discussions and which is central in metaphysical debates about personhood or the nature of agency. This is the sense in which someone with free will has, to borrow Lester's phrase, 'a spontaneous will of his own.' It seems clear to me, and to many other philosophers (for example, Descartes 1649, I, XLI, 350; Popper 1973; 1982, xx, 113-30; 1987, 145-52; Popper and Eccles 1977, 72-74, 540-42; van Inwagen 1983) that 'free will' in this sense is incompatible with determinism. This is so because an agent's free will is incompatible with that agent's actions being determined by prior circumstances, whether or not those circumstances include desires or valuations. Lester claims: '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). This is 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 if my actions are outside of my control, then I do not act freely; indeed, I do not act at all (see Hobart 1934, 346). This objection depends for its force on confusing an undetermined *act* with a random *event*. An act, specifically, an act of will, is something that is inherently under the agent's control and that is *therefore* undetermined (for a full explanation see my *forthcoming*; but the general approach is also promoted in Alvarez 2009; McCann 1974, 1975; Steward 2008, 2009; and my 2010b, sections 3 and 4).

Free will entails that an agent is capable of acting randomly and is also capable of acting irrationally, not only from weakness of will, but also purely for the hell of it. But the fact that we can to a large extent manage our interpersonal interactions and relationships shows that people do not mostly act in such ways. In large part people's behaviour is more or less predictable. There are two reasons for this. The first is that much behaviour is habitual. The second is that non-habitual actions are often informed by factual, moral and other theories that we can surmise that the agent holds, either because we share a culture with the agent, so we hold the same theories ourselves, or because we guess the agent's theories from his particular circumstances. However, the reason that people's behaviour is only more or less predictable is that it is always open to an agent to act contrary to habit, or to act contrary to the theories he holds, and even to do so rationally (see my 2010b, 26-29 and my 2013, section 3).

The great bulk of the theories any agent holds are inherited biologically or culturally. These theories concern not only how the physical world works but also how the social world works. They include theories about possible and acceptable social rules, roles and relations, about what things are valuable, what actions are obligatory, and what goals and means to them are possible, expected, acceptable or worthy. Many of the goals an agent pursues will be inherited in the sense that the

agent pursues them because he accepts a theory about his goals which has been handed down to him by his family, sub-group or the broader culture. Since these inherited theories are often vague and only partly articulated, it will sometimes be an open question for the agent, or a question he has never asked, as to whether the goal he is pursuing is either valuable for him or valuable; and this may be so even if he thinks that he ought to pursue that goal. In this way, 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. 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.

We noticed earlier that Watson deems as non-rational desires all those wants that are the products of culture or habituation rather than of the agent's rational evaluation, his independent judgement of what is good. Despite the fact that such inherited attitudes are often expressed in evaluative language, such as 'divorce is wicked,' and that they may generate feelings of guilt when we do not abide by them, Watson does not count them as values because they could come into conflict with the agent's considered judgements of value. Thus, he says, 'an agent's values consist in those principles and ends which he – in a cool and non-self-deceptive moment – articulates as definitive of the good, fulfilling, and defensible life' (1975, 215). However, Watson's policy would entail that all wants are non-rational and that no agent can have values. The reason is that a rational evaluation – an independent judgement of what is good, a considered judgement of value – cannot simply be arbitrary: it must involve argumentation; but an argument needs premises. These premises in turn cannot simply be arbitrary. So, if they are not taken from some of the agent's inherited theories, they must be products of a prior rational evaluation. But that prior evaluation will in turn need premises. Consequently, to insist that the premises of the rational evaluation must not be taken from the agent's inherited theories generates a vicious infinite regress. But if the premises of the agent's rational evaluation are taken from the agent's inherited theories, then the agent's evaluation will be deemed by Watson to be non-rational. Watson's rational evaluation which is independent of inherited theories is therefore an impossibility. Since, for Watson, an agent has values only if he does something which is in fact impossible, then, on his view, no agent has values, only non-rational desires. What this shows is that, while critical evaluation of inherited theories is necessary if we are to improve our views (of values, facts or anything else), such re-evaluation can be only a piecemeal reconstruction of part of an inherited framework, which uses some parts of that framework in criticising other parts. While we can, in principle, criticise any part of the inherited framework, we must in doing so presuppose some other parts, and thus we cannot criticise all of it at the same time (Hayek 1963, esp. 60-63; 1970, esp. 10-11, 18-22; Popper 1948, 129-32; 1994d, 134-39). Thus, an agent's inherited beliefs about what is valuable must count as his values if he is to have any values at all.

Whereas Watson deems adherence to inherited theories non-rational if those theories have not withstood the test of rational assessment, Lester deems adherence to inherited theories rational because he claims they *have* withstood the test of rational assessment.

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 mistake seems to magnify Watson's in that it not only assumes the possibility of a rational assessment independent of unexamined inherited assumptions, but it also assumes that such an assessment, in however minimal a form, has been completed even for the traditional theories that Watson deems to be held non-rationally.

Since agents necessarily have free will, and are thus free to act irrationally, (LA) is false. I am not claiming here that we have free will. I claim that, if determinism is true, then there are no agents; and that if Lester's theory of motivation were true, we would not be agents. Since agents may choose ends either contrary to their desires or for which they have no desire (in the primary sense of felt desire, which is the sense of the term that Lester favours), then Lester's postulate (a) is false and thus (LR) is also false, since it entails (a).

## 6. Self-Interest

Lester (2012, 36-39) contends that everyone is self-interested in that he pursues his own interests as he sees them (he pursues his 'self-perceived interest'). But Lester distinguishes this from psychological egoism. In order to avoid using the obscure (though popular) Kantian jargon of people being 'ends in themselves,' 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. When people act consciously, they are pursuing their second-level interests. Psychological egoism states that each person's second-level interests are concerned only with the fulfilment of his own first-level interests. Psychological altruism states that some people's second-level interests are sometimes concerned with other people's first-level interests. Lester's thesis of self-interest says that each person pursues his own second-level interests. This is consistent with psychological altruism because some people may have second-level interests in other people's first-level interests. Thus, second-level interests *belong* to the self but they may *concern* others. Obviously, there is scope for agents to have higher-level interests and to pursue those consciously, but the preceding can be taken as the basic case on which further complications are built.

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. The classification I introduced in section 4, with another category added in the light of our discussion in section 5, seems to be far more illuminating because it permits us to distinguish cases that Lester's scheme lumps together. We distinguish:

(1) what an agent desires;

- (2) what an agent thinks valuable;
- (3) what an agent thinks valuable for himself;
- (4) what an agent aims at, his goals;
- (5) what an agent thinks he has an obligation to do.

We earlier acknowledged that items in (2), (3) or (4) may also be called 'wants' or 'desires' in an attenuated sense, and that items in (1) may also be called 'values' in an attenuated sense. Since people generally think that it is often valuable to fulfil one's obligations, items in (5) may also be called 'values' and thus also 'desires.' We can acknowledge that any of those five things may also be called 'interests,' and perhaps in the same sense, since the term 'interest' seems naturally exiguous, unlike 'want' and 'desire' which seem naturally applied to appetites and are only metaphorically extended to the other cases. But 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.

For example, 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. We may distinguish the following five possibilities, amongst others:

- (A) Fred has a longing for the music he would hear at the concert and an urge to help his friend, but the urge is stronger than the longing;
- (B) as in (A) except that the longing is stronger, but Fred thinks it will be more valuable for himself to satisfy the (weaker) urge;
- (C) Fred has a longing for the music, and no appetite for helping his friend, and he thinks it will be more valuable for himself to go to the concert than to help his friend, but he thinks that his helping his friend will have a value for her that is greater than the value for him of attending the concert;
- (D) as in (C) except that Fred has no idea whether his helping his friend will have a value for her that is greater than the value for him of attending the concert;
- (E) as in (D), but Fred thinks he has an obligation to help his friend.

In (A), Fred does what he most desires. In (B), Fred does what he thinks is most valuable for himself, but not what he most desires. In (C), Fred does what he thinks is most valuable, even though it is not what he most desires and not what he thinks most valuable for himself. In (D), Fred helps his friend even though it means sacrificing an alternative which he desires more strongly and thinks is more valuable for himself, and even though he has no idea which of the two alternatives is, objectively, most valuable. In (E), Fred does what he thinks he is obliged to do, even though he does not desire to do it, and even though there is an alternative that he desires and he thinks is more valuable for him, and even though he does not know what value his chosen action has. What does Lester say about these different cases? In each case he says that Fred acted in his self-interest. Fred's helping his friend was in his self-interest 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)). What does it contribute to our understanding 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? Nothing that I can see.

It might be objected that, in cases (C), (D) and (E) it is valuable for Fred to be the kind of person who makes a sacrifice to help a friend in need or to fulfil an obligation (compare Schmidtz 1995, 111-116). Well, that is one theory; and it might be a true one. But it need not be a theory Fred holds when he makes his decision.

Fred might never have thought about whether it is valuable for him to be that kind of person; and if asked whether it is, he might sincerely answer that he does not know; or, even if he says that it is valuable for him to be that kind of person, he might have no idea whether the value for him of being that kind of person exceeds the loss in value for him of missing the concert. Whether at some deeper level, or over the long term, it *pays Fred* to make this current sacrifice is also something Fred need not have considered when making his decision, and it might be something about which he has no opinion. In fact, he might have made that decision whatever the answer to that question might turn out to be.

In response to a somewhat different complaint from Amartya Sen (1977) that self-interest theories of action are vacuous, Lester asks three rhetorical questions:

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)

But these questions, intended as rhetorical, have more or less obvious answers contrary to those assumed by Lester. 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. 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. 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). Third, we often act intentionally without motivation, in that we neither desire nor value what we do. 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. Lester's treatments of the objections of C. D. Broad and Tibor Machan (2012, 43-46) seem similarly question-begging.

Returning to Fred, we might wonder whether, in each of the five possibilities distinguished, Fred acts rationally, in the minimal sense of subjective instrumental rationality, (s). However, we cannot answer this question unless we know what Fred's aim was. It is easy to *construe* four of the five possibilities as instrumentally rational behaviour by attributing to Fred an appropriate aim, thus: in (A), to do what he most desires; in (B), to do what is most valuable for himself; in (C), to do what is most valuable; in (E), to do what he is obliged to do. In (D), we might postulate that Fred thinks he is obliged to help his friend; but that would reduce the case to (E). If we maintain (D) as a separate case, it seems that the only aim we could attribute to Fred to render his action instrumentally rational is the aim of helping his friend. But that raises the question of whether that aim is rational, given the absence of desire, higher valuation and assumed obligation. We might say that, while Fred's action was locally rational (relative to his immediate aim), it is doubtful that it was globally rational (relative to the aims that Fred regards as most important). If the question occurs to Fred, it may prompt him to undertake the sort of piecemeal reappraisal of his ends discussed in section 5, above.

## 7. Maximisation

According to Lester, as agents, we seek to maximise our want-satisfaction in the sense that, '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.’ This is what it means to feel the greatest desire for the option finally chosen (2012, 50-51).

There are a number of problems with this view, including the following two, which were pointed out by Herbert Simon. First, most of our decisions are habitual or conventional rather than reasoned (Simon 1997, 102-103, 107-109), so what is chosen is the usual rather than the best or the most want-satisfying. 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 (1997, 73-75, 77, 93-97), so we often cannot identify the best, or the most want-satisfying. Simon called this ‘bounded rationality’ and he insisted that, as a consequence, decision-makers generally ‘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 (1997, 118-129). In option appraisals, managers in business and in the public-sector typically attempt to construct very simplified pictures of reality which seem likely to enable them to identify, and thus rule out, the more disastrous options. The choice amongst the options that remain is then often rather arbitrary (Simon 1997, 264-65; Peters and Waterman 1982, 29-54).

Lester retorts to Simon, in a footnote, 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 (2012, note 63, 210). This is a familiar response to Simon; but it seems to be a wholly inadequate one. We saw this in connection with habitual and conventional actions in section 5; but it is also so with regard to option appraisals, for the following reasons. *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. 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. *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. But that would be to concede Simon’s point, because to decide arbitrarily to search no farther *just is* to decide to satisfice rather than to maximise. *Third*, if Lester means, not that the decision-maker decides arbitrarily to search no more, but that he calculates whether further search is worthwhile, then we get a vicious circle. Call the options for solving the decision-maker’s problem ‘first-level options.’ Call the options for gathering information about first-level options ‘second-level options.’ The decision-maker’s situation is that he does not have enough information about the first-level options before him to be able to say that one of them is better than the rest, and he does not know whether there are additional first-level options that would be better than those he has before him. To decide whether to search for more information, he must, on Lester’s view, evaluate second-level options. But he cannot know what all the second-level options are, since there may be (and doubtless are) sources of information he does not know about. Even for the second-level options he can identify, he cannot know in advance what all the search costs will be. And, crucially, he has no way of identifying the potential benefits of additional information without knowing what that information will tell him about the costs and benefits of the first-level options he is investigating (that is, those options he is investigating precisely in order to *discover* their costs and benefits). To know the benefits of the second-level options, he must know the costs and benefits of the first-level options.

So, on Lester's view, the decision-maker has to solve the second-level problem before he can solve the first-level problem; but he has to solve the first-level problem before he can solve the second-level problem.

Since agents are often unable to identify an option which would maximise the achievement of their ends, then, according to Lester's postulate (b), they are often unable to identify a suitable means to their end, which means that, according to (LR), they often fail to exhibit instrumental rationality in their actions. That is inconsistent with (LA).

## 8. Morals

We have seen that Lester thinks that every agent is self-interested in that he is bound to want only things that he values; but Lester contends that some of the things an agent values may be altruistically or morally valued (2012, 41). For Lester, recall, to value something is to desire it, so he attributes to agents moral desires or 'sentiments' (2012, 51). Lester views morality as impartial in that moral principles are not concerned with the interests of a specific agent, but partial in that moral principles must motivate agents who sincerely hold them (2012, 39-40):

*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).

Lester's position here seems to fly in the face of human experience. I seriously doubt 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. One type of case involves weakness of will, 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. The varied situations and ways in which this happens has been explored extensively in literature, and also in film, and its effects have been analysed in some psychological studies of cognitive dissonance (see Tavis and Aronson 2007). Typically, the person who knowingly does wrong 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. He tries to delude himself into thinking that he has not really done wrong 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.

Lester refers to Hare for further argument for his position. The view of Hare (1952) is that the primary function of words such as 'good' and 'right' is to commend or prescribe, but that they also have a secondary function to indicate the presence of those non-value properties the presence of which inclines us to say that something is good or right. On this view, there is no property of goodness itself or rightness itself; there are just non-value properties and acts of prescription. Since moral judgements are universal, one's moral prescriptions for others are also moral prescriptions for oneself if one is in the same circumstances. Thus, someone contemplating an action which he sincerely says (or thinks) is immoral, commands himself not to perform that action, and thus he does not perform it. One might question why someone may not

do something that he simultaneously commands himself not to do, but the objection I want to raise here is that we can discuss dispassionately whether a thing is good or right, without making any prescriptions; indeed, we can even prescribe what we think is not good or not 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). However, Hare is here merely re-describing the examples to fit his theory. 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. 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. 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); and similarly if he says 'the right thing to do is the wrong thing to do.' 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. 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. Hare's view is mistaken because it omits mention of the properties of goodness and rightness.

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?' His assumption, of course, is that the answer is 'no.' However, my answer is 'yes.' I can recall many examples of my doing things that I clearly thought were immoral. Anyone else could do the same, provided he has not deluded himself with the self-justifying chicaneries of cognitive dissonance.

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*...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).

Is it possible to have a *feeling* with so reticulated a propositional content as *no one should ever do an action of type A in these circumstances* (where an appropriate action-description is substituted for 'A') or *lying is usually immoral* 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 supposition seems ridiculous. 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. 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, but the feeling itself does not have the structured complexity of the thought. Lester seems driven to supposing that it does, not simply by his mechanistic theory

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

## 9. Conclusion

Lester articulates and defends two hypotheses: that all agents are instrumentally rational whenever they act (LA); and that an agent is instrumentally rational if and only if he seeks to maximise the satisfaction of his own wants (LR). We have seen that (LA) is false because it cannot accommodate the phenomenon of weakness of will and it involves a passive conception of agency which leaves no place for the agent. We noted that (LR) can be analysed as a conjunction of a minimal conception of instrumental rationality

(s) an action is (subjectively) instrumentally rational if and only if the agent thinks it is a suitable means to his subjective ends

and two postulates

(a) an agent's subjective ends are always the ends of his desires;

(b) a means to an end is suitable if and only if it maximises the achievement of that end.

We have seen that postulate (a) is false because agents often have ends that they value but do not desire (in the primary sense of felt desire) and because they often have ends that they neither value nor desire; and that postulate (b) is unacceptable since agents are often unable to identify an option which would maximise the achievement of their ends. This entails that (s) is a better account of instrumental rationality than (LR) and that, if (LR) were true, then (LA) could not be true.

Lester also defends two other theses: that every agent always acts out of self-interest and that no agent can do what he genuinely feels to be immoral. We have seen that the first thesis can be maintained, but only in a sense in which it is quite vacuous; and that the second thesis is false.

Lester says that we cannot understand people's behaviour unless they are instrumentally rational (2012, 15, 16). That is true; but we need only the minimal conception of instrumental rationality, (s), to make sense of intentional action. The fact that people are often instrumentally irrational means that people are often not predictable; but this does not threaten theoretical social science because its generalisations are either explicitly statistical or their truth requires only that the vast majority of agents act instrumentally rationally in response to a given change (see my 2013, section 4).

At a number of places Lester exhibits a concern to avoid giving encouragement to paternalism. When he admits that people can be mistaken about their own interests, he explains why he thinks this does not provide an argument for paternalism (2012, 13). It also seems that he regards acknowledging the possibility of weakness of will as opening the door to coercion of people for their own good (2012, 19, 27, 33-34). However, my argument should not be taken to be making a case for paternalism. It is true that a person may be mistaken about his own objective ends; but it is true *a fortiori* that other people can be mistaken about that person's objective ends. A central task of any agent's life is to discover his own objective ends through experiments in living; and this requires freedom. Similarly, an agent is better placed than an observer to recognise whether an action of his is due to weakness of will or to personal idiosyncrasy; and if he wants assistance in overcoming weakness of will, that is his call to make. Finally, that people often do

what they think to be wrong indicates that there is a propensity to evil in each of us, so that the most beneficial social arrangement will be one that disperses power and gives quick feedback, positive or negative, on actions performed, which is what open markets do and what overbearing government frustrates.

I have argued that Lester's *homo economicus* accounts of rationality and agency are false. This is not to say that they are worthless. In economics, *homo economicus* can be defended as a simplifying assumption which has illuminating consequences which enable us to make predictions. Such simplified models form part of every science (Popper 1994b, 162-70). In contrast, an argument for freedom requires a realistic conception of the agent, because it must show that freedom is valuable for real agents. It does not seem to be much use showing that freedom would be great for *homo economicus*, if real people are not much like *homo economicus*. To defend freedom we need to show that it would be good for real people with limited rationality who behave largely habitually, who have values and obligations as well as desires, and aims which may be none of those, who learn by modifying inherited theories, who often act irrationally and also in ways they think to be immoral, who are remarkably ignorant and fallible, and who often behave unpredictably.

## References

- Alvarez, Maria. 2009. 'Actions, Thought-Experiments and the "Principle of Alternate Possibilities.'" *Australasian Journal of Philosophy* 87/1: 61–81.
- Descartes, Rene. 1649. 'The Passions of the Soul.' In his *Philosophical Works*, Vol. 1, trans. E. S. Haldane & G. R. T. Ross, corrected edition (329-427). Cambridge: Cambridge University Press (1931).
- Frankfurt, Harry. 1971. 'Freedom of the Will and the Concept of a Person.' *The Journal of Philosophy* 68 (1): 5-20.
- Frederick, Danny. 2010a. 'Unmotivated Intentional Action.' *Philosophical Frontiers* 5 (1): 21-30. Available here: [http://independent.academia.edu/DannyFrederick/Papers/311267/Unmotivated\\_Intentional\\_Action](http://independent.academia.edu/DannyFrederick/Papers/311267/Unmotivated_Intentional_Action)
- Frederick, Danny. 2010b. 'Popper and Free Will.' *Studia Philosophica Estonica* 3.1: 21-38. Available here: <http://www.spe.ut.ee/ojs-2.2.2/index.php/spe/article/view/93>
- Frederick, Danny. 2013. 'Popper, Rationality and the Possibility of Social Science.' *THEORIA* 28 (1): 61-75. Available here: <http://www.ehu.es/ojs/index.php/THEORIA/article/view/1879>
- Frederick, Danny. *Forthcoming*. 'Free Will and Probability.' *Canadian Journal of Philosophy*.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Hayek, Friedrich. 1963. 'Rules, Perception and Intelligibility.' In his *Studies in Philosophy, Politics and Economics* (43-65). London: Routledge and Kegan Paul (1967).
- Hayek, Friedrich. 1970. 'The Errors of Constructivism.' In his *New Studies in Philosophy, Politics, Economics and the History of Ideas* (3-22). London: Routledge and Kegan Paul (1978).
- Hobart, R. E. 1934. 'Free Will as Involving Determination and Inconceivable Without it.' In P. van Inwagen & D. W. Zimmerman (eds.), *Metaphysics: the Big Questions* (343-55). Oxford: Blackwell (1998).
- Lester, Jan. 2012. *Escape from Leviathan: Libertarianism without Justificationism*, paperback edition. Buckingham: University of Buckingham Press.
- McCann, Hugh. 1974. 'Volition and Basic Action.' In his 1998 (75-93).

- McCann, Hugh. 1975. 'Trying, Paralysis and Volition.' In his 1998 (94-109).
- McCann, Hugh. 1998. *The Works of Agency*. Ithaca: Cornell University Press.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.
- Peters, Tom and Waterman, Robert. 1982. *In Search of Excellence*. London: Harper and Row.
- Popper, Karl. 1948. 'Towards a Rational Theory of Tradition.' In his *Conjectures and Refutations: the Growth of Scientific Knowledge*, fourth (revised) edition (120-135). London: Routledge and Kegan Paul (1972).
- Popper, Karl. 1973. 'Of Clouds and Clocks.' In his *Objective Knowledge*, corrected edition (206-55). Oxford: Oxford University Press.
- Popper, Karl. 1982. *The Open Universe*. London: Routledge.
- Popper, Karl. 1987. 'Natural Selection and the Emergence of Mind.' In G. Radnitzky & W. W. Bartley (eds.), *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge* (139-55). La Salle, IL: Open Court.
- Popper, Karl and Eccles, John. 1977. *The Self and its Brain*. London: Springer-Verlag.
- Popper, Karl. 1994a. 'The Myth of the Framework.' In his 1994c (33-64).
- Popper, Karl. 1994b. 'Models, Instruments, and Truth.' In his 1994c (154-84).
- Popper, Karl. 1994c. *The Myth of the Framework*. London: Routledge.
- Popper, Karl. 1994d. *Knowledge and the Body-Mind Problem*. London: Routledge.
- Schmidtz, David. 1995. *Rational Choice and Moral Agency*. Princeton: Princeton University Press.
- Sen, Amartya. 1977. 'Rational Fools.' *Philosophy and Public Affairs* 6 (4):317-44.
- Simon, Herbert. 1997. *Administrative Behavior*, fourth edition. New York: The Free Press.
- Steward, Helen. 2008. 'Moral Responsibility and the Irrelevance of Physics.' *Journal of Ethics* 12/2: 129-145.
- Steward, Helen. 2009. 'Fairness, Agency and the Flicker of Freedom.' *Nous* 43/1: 64-93.
- Tavris, Carol and Aronson, Elliot. (2007). *Mistakes Were Made (But Not By Me)*. New York, Harcourt.
- Tolstoy, Leo. 1911. *The Devil*. In his *The Devil and Other Short Stories*, translated by Louise and Aylmer Maude, revised by Richard Gustafson, 191-236. Oxford: Oxford University Press (2003).
- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- Watson, Gary. 1975. 'Free Agency.' *Journal of Philosophy* 72 (8): 205-220.