Wlodek Rabinowicz

Safeguards of a disunified mind

Article (Accepted version) (Refereed)

Published in *Inquiry* 57, 2014, 365-383

Safeguards of a Disunified Mind*

Wlodek Rabinowicz

This paper targets 'pragmatic' arguments for various rationality requirements on agents' beliefs or preferences. An argument of this kind focuses either on the benefits of satisfying such requirements, or – more often – on the costs of their violation. It is the arguments of this second type that I am interested in. For a given requirement, an argument of this kind typically takes the form of a proof that whoever violates the requirement in question is vulnerable to exploitation: She can be taken for a ride by a clever bookie, even if the latter doesn't know more than she does herself. The general idea of this way of approaching the costs of violations is due to Frank Ramsey:

> If anyone's mental condition violated these laws [= the laws of probability], … [h]e could have a book made against him by a cunning bettor and would then stand to lose in any event. (Ramsey 1990 (1926), p. 78)

Pragmatic arguments make use of different exploitation set-ups: (i) synchronic Dutch Books, for the violations of the standard probability axioms, (ii) diachronic Dutch Books, for the violations of diachronic probability principles, such as Reflection and Conditionalization, and (iii) Money Pumps, for the violations of the acyclicity requirement on preferences.

When we examine these various examples, one thing stands out: The different exploitation set-ups are based on the same underlying assumption. Thus, consider an agent who is logically and mathematically competent, but violates a given rationality

* This paper is a significantly revised version of Rabinowicz (2012), which in turn was a revised short version of Rabinowicz (2008). I am indebted to the Swedish Collegium of Advanced Study (SCAS) in Uppsala which provided me with an excellent work environment. The paper was prepared for a workshop in diachronic rationality at the Centre for the Study of Mind in Nature (CSMN) in Oslo (August 2013). It was also presented at departmental philosophy seminars in Leeds (April 2013) and Glasgow (October 2013), at the Eighth Conference on Moral and Political Philosophy in Sophenberg, Denmark (June 2013) and at the Polish-Scottish Conference in Epistemology and Political Philosophy in Warsaw (September 2013). I am grateful to the organizers of these events, and to the participants for their helpful and insightful comments and suggestions.

requirement. (This presupposes that the requirement itself is not a purely logical or mathematical demand.) Also, suppose she prefers to be better off rather than worse off and acts accordingly. Then, as we shall see, even if such an agent violates the rationality requirement under consideration, she cannot be exploited if she makes her decisions in a *unified* fashion. To be exploited she has to be disunified in her decision-making, i.e., to make decisions on various issues she faces one by one, instead of addressing them together. A disunified agent decides on each component in a package separately, rather than jointly. The different exploitation set-ups are all based on the assumption of disunification.

An agent can be disunified synchronically or diachronically. In the synchronic case, there is a time when she is presented with several opportunities, each of which she can accept or reject. If she is disunified, she deals with each opportunity separately and on its own merits. A unified decision-maker would in sucha case instead make a single choice of a particular *configuration* of the opportunities. In the diachronic case, different opportunities are offered at different times. Suppose the schedule of offers is known beforehand by the agent. She is diachronically disunified if she defers her decisions concerning later offers to the times at which the offers will be made and need to be responded to. A unified approach would again involve a single choice of a particular configuration of opportunities, present and future. [1] It would thus amount to a choice of a particular branch in the agent's decision tree – a branch that extends

[1] For an extended argument in favor of unified decision making as a shield against exploitation, see McClennen (1990), who discusses diachronic unification under the label of "resolute choice". More precisely, resolute choice appears to be a somewhat restricted and possibly more realistic form of diachronic unification. A resolute agent does not make an *irrevocable* choice of a particular configuration of present and future opportunities. Instead, she forms a plan as to what to choose at different future occasions. Embarking upon such a plan of action is a factor that significantly modifies the context of her future decisions, but there's still a possibility for a resolute agent to deviate from what she has planned to do. However, being resolute, she does not deviate. Why she doesn't do it even in those cases in which she might be expected to prefer to do so is another matter. It might be because her previously adopted plan modifies her preferences at later occasions so as to put them in line with the plan adopted, but it might also be because her commitment to follow the plan takes precedence over her preferences. For his own part, McClennen favours the second avenue, if I understand him correctly, but he rejects the suggestion that the mere adoption of a plan by the agent's earlier self obligates her later selves to compliance (cf. ibid., fn 12 to chapter 9, p. 285). Instead, he favours the interpretation of resolute choice on which compliance is justified by the plan being a *reasonable compromise* between the preferences of the agent's earlier and later selves – a compromise from which all selves involved in the implementation of the plan can profit (cf. ibid., section 12.6.).

For a broader picture of the various aspects of planning and diachronic self-governance, see Bratman (1987), (2007), (2012).

from the root of the tree to its very top.[2] A disunified agent, by contrast, would instead make separate choices at each node of the tree, with each choice concerning only the immediate move at the node in question. (As we shall see, this doesn't presuppose myopia. A disunified agent's choice of a move at hand might well take into consideration the predictions she makes about her future moves.)

Unification in decision making requires consideration of complex choice alternatives. For various reasons, we might often find it easier to deal with different issues one by one, rather than in a wholesale manner. Furthermore, in diachronic cases, pre-commitments might be difficult or even impossible: An agent's self-control might not reach far enough into the future. All this imposes significant practical limitations on unified decision making.

Since the exploitation set-ups only work for disunified agents, pragmatic arguments for various constraints on beliefs and preferences only support *conditional* recommendations: "You should satisfy this constraint if you are going to make your decisions in a disunified manner." In other words, pragmatic arguments identify constraints that function as safeguards of a disunified mind – that decrease the costs of disunity in decision-making. Obeying such constraints shields the disunified agent from potential exploitation.

I am not going to provide a conclusive defense of this interpretation of pragmatic arguments. I will, however, support it by several examples that illustrate the intimate connection between exploitability and disunification.

I will also argue against a popular view that diachronic exploitation set-ups can be avoided by disunified agents who have foresight. On the view I am going to criticize, disunified agents are not vulnerable to exploitation if they know what's kept in store for them, i.e., if they know, beforehand, the schedule of offers they are going to receive. I will show that foresight is not enough.

Note that on the view I defend diachronic pragmatic arguments are *more compelling* than synchronic ones. The reason is that unified decision making is much more difficult to manage diachronically than synchronically. Consequently, there are

[2] The choice of a branch in a decision tree amounts to a choice of a specific *plan* of action. If a unified agent envisages a possibility that she might at some point deviate from the chosen plan, she is well-advised to do more than just form a plan. She might instead choose a *strategy*: an assignment of moves to each node in the decision tree. As is easily seen, a strategy determines a plan of action (= a sequence of consecutive moves in which every step follows the strategy), but it does more than this: It specifies the moves to be made even at the nodes that are unreachable if one follows the strategy at the earlier nodes.

stronger reasons for an agent to satisfy the constraints that would decrease the costs of diachronic disunification.

Isaac Levi (2002) has a very different view of the status of pragmatic arguments. According to him, only synchronic arguments have a good claim to validity. In his (2006), he modifies this view and takes a position that seems to deny validity to all pragmatic arguments, whether diachronic or synchronic. Before I explain why he takes these views and why I think he is mistaken, I need to present some examples of the arguments of both kinds, in order to provide the background.

1. A synchronic Dutch-book argument for probability laws

In this argument, it is assumed that an agent's probability assignments – her degrees of belief - are her guides to action. As such, they are embodied in her betting dispositions, or betting commitments.[3] More precisely, the agent's probabilities are given by her *betting rates*.

[3] This idea, which was most famously defended by Ramsey and de Finetti, can be traced at least as far back as to Kant's "Transzendentale Methodenlehre", towards the end of the *Kritik der Reinen Vernunft.* On Kant's view, the strength of belief can be measured by the agent's willingness to bet. An agent who is willing to bet a ducat on a proposition might balk at betting ten ducats, not to speak of betting "the happiness of the whole life":

> Der gewöhnliche Probirstein: ob etwas bloße Überredung, oder wenigstens subjective Überzeugung, d.i. festes Glauben sei, was jemand behauptet, ist das *Wetten*. Öfters spricht jemand seine Sätze mit so zuversichtlichem und unlenkbarem Trotze aus, daß er alle Besorgnis des Irrtums gänzlich abgelegt zu haben scheint. Eine Wette macht ihn stutzig. Bisweilen zeigt sich, daß er zwar Überredung genug, die auf einen Dukaten an Wert geschätzt werden kann, aber nicht auf zehn, besitze. Denn den ersten wagt er noch wohl, aber bei zehnen wird er allererst inne, was er vorher nicht bemerkte, daß es nämlich doch wohl möglich sei, er habe sich geirrt. Wenn man sich in Gedanken vorstellt, man solle worauf das Glück des ganzen Lebens verwetten, so schwindet unser triumphierendes Urteil gar sehr, wir werden überaus schüchtern und entdecken so allererst, daß unser Glaube so weit nicht zulange. So hat der pragmatische Glaube nur einen Grad, der nach Verschiedenheit des Interesse, das dabei im Spiele ist, groß oder auch klein sein kann. (A824-5, B852-3).

A reference to this passage has recently been made by Chignell (2007), p. 333, and Blamey (2011), pp. 79 and 189. For a short discussion of Kant's use of betting as a touchstone ("Probirstein") of the strength of belief, and in particular of the difference between his betting arrangement and the kind of arrangement presented in what follows, see Eriksson & Rabinowicz (2012). In that paper, we suggest that Kant probably considers bets at even odds, in which one wagers a monetary amount against an equal monetary amount wagered by one's opponent. The agent's willingness to bet decreases when the monetary amount increases because of the diminishing marginal uility of money: as more money is being wagered, the potential win is worth less and less in utility terms as compared with the potential loss. In this respect, Kant's betting arrangement sharply differs from the modern set-up which I describe below. In the latter, one implicitly assumes that utility is linear with money, i.e. that the marginal utility of money is constant.

As a curiosity, it might be mentioned that Kant declared himself prepared to bet all his possessions on the proposition that at least some of the (other) planets we see are inhabited (provided such a bet could be empirically resolved):

> Wenn es möglich wäre durch irgendeine Erfahrung auszumachen, so möchte ich wohl alles das Meinige darauf verwetten, daß es wenigstens in irgendeinem von den Planeten, die wir sehen, Einwohner gebe. Daher sage ich, ist es nicht bloß Meinung, sondern ein starker Glaube (auf dessen Richtigkeit ich schon viele Vorteile des Lebens wagen würde), daß es auch Bewohner anderer Welten gebe. (A825, B853)

To see what this means, consider a bet on a proposition *A*. Assume that the bet costs *C* and pays *S* if won, where *S* and *C* – the *price* and the *stake* of the bet, respectively – are monetary amounts.[4] A bet is *fair* for an agent if the latter is prepared to take each of its sides: either to buy it or to sell it. Assume that some bets on *A* indeed are fair for a given agent (which isn't obvious) and, furthermore, that they all exhibit the same ratio between their prices and stakes: If the price increases, the stake must increase in the same proportion for the bet to remain fair. The constant price-stake ratio for fair bets on *A* is the agent's *betting rate* for *A*.

The assumption that bets remain fair under proportional increases of prices and stakes is certainly problematic as a general claim, but it is arguably reasonably realistic within a limited range in which the monetary amounts *S* and *C* are not too high and not too low. The explanation is that, within this range, utility is linear with money, which means that proportions between losses and gains in money tend to be equal to proportions between corresponding losses and gains in utility.

On the betting interpretation, the agent's *probability* for *A*, *P*(*A*), is identified with her betting rate for *A*. The higher the price the agent is willing to pay for a bet on *A* with a given stake (or the higher the price she demands for selling such a bet), the higher is her probability for *A*.[5]

Note that, on this interpretation of probabilities as betting rates, the expected monetary value of buying a fair bet on *A* with price *C* and stake *S* is zero:

$[P(A) \times S] - C = [C/S \times S] - C = 0.$

Similarly, selling such a bet has the expected value zero:

$C - [P(A) \times S] = C - [C/S \times S] = 0.$

This might be thought to explain why the agent is equally willing to buy a fair bet as to sell it. On the assumption that, for moderate monetary amounts, utility is linear with money, such an explanation is reasonable for agents who are expected-utility maximizers. But in the present context, in which we consider agents who violate

---

[4] Betting terminology varies. Sometimes, the term "stake" refers to what the agent stakes or wagers and what she would lose if she lost the bet.. While in the usage I favour a stake instead is what is "at stake"– it is the amount of money that can be won.

[5] Cf. de Finetti (1990), p. 75: "The probability *P*(*E*) that You attribute to an event *E* is therefore the certain gain *p* which You judge equivalent to a unit gain conditional on the occurrence of *E*: in order to express it in a dimensionally correct way, it is preferable to take *pS* equivalent to *S* conditional on *E*, where *S* is any amount whatsoever, one Lira or one million, $20 or £75."

In other words, if *P*(*E*) – the subjective probability you attribute to *E* – equals *p*, then *pS* is your fair price for a bet on *E* with a stake *S*: You judge it "equivalent to *S* conditional on *E*". Note that if we let $C = pS$ and if we assume that $S > 0$, then it immediately follows that $p = C/S$.

various rationality constraints, an explanation on the expected-utility lines might be unavailable. If an agent violates some of the basic assumptions of the expected-utility theory (such as the standard probability axioms), then we cannot explain her behaviour by an appeal to expected-utility considerations. The identification of the agent's probabilities with her betting rates is therefore something of a problem in the context of rationality violations. This somewhat undermines the pragmatic arguments that focus on the violations of various synchronic or diachronic probability laws, since such arguments *presuppose* that probabilities are cashed out as betting rates.[6] I shall, however, disregard this issue in what follows.

There is one thing we should note before we continue. An adherent of the betting interpretation takes probabilities, i.e., degrees of belief, to be betting rates. Does this mean that he also takes beliefs to be betting dispositions? Not necessarily. For the pragmatic arguments to work, it is not necessary to assume that an agent's beliefs are *identical* with her betting dispositions (or betting commitments). What one needs to assume is only that the latter are in line with the former, i.e., that the agent is disposed (or committed) to bet in accordance with her beliefs. This is enough to guarantee that degrees of belief will at the same time be measures of the agent's dispositions (or commitments) to bet.

Let us move on. A *Dutch book* is a system of bets on various propositions which is such that, if an agent accepts all those bets, she must suffer a loss come what may, i.e., whatever turns out to be the case. A Dutch book is *synchronic* or *diachronic* depending on whether the bets in the book are offered at the same time or at different occasions.

If probabilities are betting rates, then the agent who violates the standard probability laws is vulnerable to a synchronic Dutch book. This provides a pragmatic argument for obeying the laws in question.[7]

As an example, consider the addition axiom for probabilities,

$P(A \vee B) = P(A) + P(B)$, if A is logically incompatible with B.

[6] The identification of probabilities with betting rates is problematic in other contexts as well (cf. Eriksson and Rabinowicz 2012). But it creates special problems in the present context.

[7] Actually, this argument also has a positive part: It can be shown that an agent whose betting rates satisfy the standard probability axioms is not vulnerable to a Dutch book. (More precisely, she is invulnerable to synchronic Dutch books; to avoid diachronic ones she must also satisfy such principles as Reflection and Conditionalization.) In what follows, however, I focus on the negative part of the argument.

Suppose your probability assignments P violate this axiom. For example, suppose that for some logically incompatible A and B, $P(A \veebar B) < P(A) + P(B)$. In such a case, I can offer you a bet on A and another bet on B, each with the same stake $S > 0$, and with prices $P(A)S$ and $P(B)S$, respectively. At the same time I can ask you to *sell* a bet on $A \vee B$, again with the same stake S and a price $P(A \vee B)S$. Given your probabilities, all these bets are fair. Their price-stake ratios equal your probabilities for the propositions on which the bets are to be made: $P(A)$, $P(B)$ and $P(A \vee B)$, respectively. But if you accept the three bet offers, you are bound to make a net loss. Your loss will be $P(A)S + P(B)S - P(A \vee B)S$, which is the price difference between what you need to pay for the first two bets and what you receive for the bet you sell. Since $P(A \vee B) < P(A) + P(B)$, you need to pay more than you receive. As for the stakes, if either A or B turns out to be the case, you are going to win one of the bets you have bought (just one, since A and B are incompatible), but lose the one you have sold. Thus, you will receive S, but at the same time you will have to pay back the same amount. If neither A nor B turns out to be the case, no bets will be won by either party. Thus, no stake payments will be made. This means that, whatever happens, you net loss will equal $(P(A) + P(B) - P(A \vee B))S$. [8]

As an aside, an important limitation of this argument needs to be mentioned at this point. Even a fully rational agent can be exploited by parties who know more than the she does. There is nothing strange in this: knowledge is power. In pragmatic arguments, if such arguments are to have any bite, it is therefore essential that the exploiter doesn't know more than the agent herself. But since the exploiter needs to know the agent's probability assignments in order to set up a Dutch book, it follows that the agent must be assumed to know her probability assignments as well. Needless to say, this assumption of self-knowledge on the part of the agent might well be questioned: knowing the strength of one's beliefs is not atrivial matter.[9] In what follows, however, I am going to disregard this problem.

[8] If the sum of the agent's probabilities for A and B instead is *lower* than her probability for the disjunction $A \vee B$, then the exploitation set-up is reversed: the agent is asked to sell the bets on A and on B and to buy the bet on $A \vee B$. No exploitation scheme is available only if $P(A \vee B) = P(A) + P(B)$, as required by the addition axiom.

[9] This point, as we have seen, was already made by Kant.

What I instead want to focus on is another limitation of the argument: the role played by disunification. A violator of the addition axiom is being exploited in this set-up because her decision-making is disunified: She decides on each bet separately, rather than jointly. If she did the latter, then – assuming she is logically and mathematically competent – she would certainly not choose to accept the whole bet package: She would recognize that at least one combination of actions – refusing all the bets – would be better for her whatever happens. In this unified mode, she might however still decide to accept one or two bet offers, say, to sell the bet on the disjunction $A \vee B$ but refuse to buy the bets on A and on B. Obviously, this would not give her a guaranteed loss.

At one point, Skyrms (1980) went further than this and suggested that an agent who is vulnerable to a synchronic Dutch book must be logically confused (and not just disunified in her decisions). Such an agent evaluates one and the same betting arrangement differently depending on the way it is presented to her: as a collection of three fair bets, or as one composite opportunity that would give her a loss whatever happens. It is only if viewed in the latter way that she evaluates the arrangement as bad. However, this suggestion of logical confusion is unconvincing. Admittedly, the agent we consider does not view any of the bets in the package as unattractive, but she might well assign a negative value to the bet package as a whole. What this means, howevr, is only that her evaluations aren't additive: The value she ascribes to the whole differs from the sum of the values she ascribes to its parts. Non-additivity in evaluation need not be questionable as such and it certainly doesn't presuppose a logical mistake of any sort (cf. Schick 1986). That instrumental value need not be additive is obvious; in economics this feature is referred to as complementarity. A knife is more useful than a fork, but a knife and a fork together are more useful than two knives. That final value (value in itself, for its own sake) need not be additive either is less obvious but its non-additivity has been recognized by a fair number of philosophers starting at least with G. E. Moore (1903). The latter referred to this phenomenon as "the principle of organic unities." Admittedly, the 'expectational' value possessed by bets is not final and it does not seem to be instrumental either, at least not in the literal sense, but it is by no means obvious that such value must satisfy additivity.

We need to say something about the case in which the agent violates the probability axiom that requires logical truths to be assigned probability one. If her probability for a logical truth *A* is higher than one, she is willing to buy a bet on *A* for a price that is higher than the stake to be won. If it is lower than one, she is willing to sell a bet on *A* for a price that is lower than the stake she is guaranteed to lose. Thus, in each case, she is exploitable. Since this exploitation set-up consists of just one bet, it does *not*, in order to work, require the agent to be disunified in her decision making. But it only works if the agent *is* logically confused (does not recognize that *A* is logical truth) or if she doesn't mind making losses. This means that the case under consideration is consistent with the claim we are defending: Disunification is a necessary pre-condition of exploitability for a logically and mathematically competent agent, who prefers to be better off than worse off.

2. A diachronic Dutch-book argument for Reflection

The Principle of Reflection stipulates that one's current conditional probabilities should reflect one's hypothetical future probabilities. More precisely, letting *P* be the agent's probability at time *t*, and *P'* her probability at $t' \geq t$,

Reflection: $P(A/P'(A) \leq k) \leq k$, provided that $P(P'(A) \leq k) > 0$.

The analogous condition applies if we replace all the occurrences of ≤ by ≥, or by =, in the principle above.

The intuitive plausibility of Reflection as a general constraint might well be questioned: Just think of cases in which we have good grounds to expect a degeneration of our epistemic capacities that will unduly draw down (or draw up) our future probabilities for *A*. In such cases, obeying Reflection seems clearly unjustified. Nevertheless, as has been shown by van Fraassen (1984), an agent whose probability assignments violate Reflection, for whatever reason, is vulnerable to a diachronic Dutch Book. Instead of presenting his proof in full generality, let us consider an example, due to Christensen (1991), which illustrates this point. Suppose an agent's probability assignment *P* at *t* fails to reflect her hypothetical probabilities *P´*at *t'*: At *t*, she suspects that at *t'* her probability estimate of *A* will be too low. In particular, therefore,

(i) $P(A/P'(A) \leq ½) = 3/4$.

Letting E stand for $P'(A) \leq ½$, suppose that

(ii) $P(E) = 1/5$.

At *t*, a bookie offers the agent two bets:

(1) a bet on *E*, with price 1 and stake 5;

(2) a bet on *A conditional* on *E*, with price 15 and stake 20.

In a conditional bet, if the condition turns out to be false, the bet is called off and its purchase is refunded. On the betting interpretation, conditional probabilities equal betting rates for conditional bets. Therefore, given our assumptions about the agent's probabilities, it is easy to see that bets 1 and 2 are fair: their price-stake ratios are 1/5 and 3/4, respectively.

Then, at *t', if E* turns out to be true, but not otherwise, the bookmaker offers to *buy* from the agent a third bet:

(3) a bet on *A*, with price 10 and stake 20.

If *E* is true, then at *t'* the agent should be willing to sell this bet. We know that, if *E* is true, the agent's probability for *A* at *t'* will not exceed ½.

If the agent accepts all the bet offers (including the third one, if that offer is made as well), she will lose 1 unit whatever happens. If *E* is false, she will lose 1 unit on her bet on *E*, while the conditional bet on *A* given *E* will be called off and no bet offer will be made at *t'*. If *E* is true, she will win the bet on *E* and her conditional bet on *A* will be on. But then, at *t*', the bookie will buy back this bet on *A at a lower price* (bet 3). Since the price difference (15 – 10 = 5) exceeds by 1 unit her net gain from the bet on *E* (5 – 1 = 4), the agent will again suffer a total loss.

There is an obvious objection to this line of reasoning. A pragmatic argument for a rationality constraint is supposed to demonstrate that a violation of this constraint would lead to a guaranteed loss by the violator's own lights. To be effective, such an argument should therefore be based on the assumption that the agent to be exploited knows at least as much as her would-be exploiter. We have already pointed this out before. But then, in the diachronic case, the agent must know, beforehand, the what bets are kept in store for her. Which means she must know that she is being taken for a ride. In other words, in the case of diachronic set-ups, the agent must have *foresight.* But, the objection continues, an agent with foresight will surely upset the bookie's

evil design by simply refusing to accept the earlier bets in the book. Thereby, the whole book will crumble: By refusing bets at *t*, the agent prevents the bet offer at *t'* – an offer which, if it were made (i.e. if *E* would turn out to be the case), she would be willing to accept, but the prospect of which she now, at *t*, finds unattractive.[10] (For this line of reasoning, cf. Levi 1988, and Maher 1992.)

Skyrms (1993) shows how this objection can be disarmed.[11] Suppose the bookie is *persistent*. Persistency means that the later bet offers in the exploitation scheme are not conditioned on the acceptance of the earlier ones. Thus, assuming that the agent knows the set-up, including the persistency of the bookie, she knows that the latter is bound to offer to buy bet 3 at *t'* if *E* will turn out to be the case – that he will do it even if the agent at *t* were to refuse the bets offered at that time. Suppose also that the bookie makes all the three bets 'more than fair': For each bet she accepts, the agent will get a small reward ε. Still, the reward is so small that $3\varepsilon < 1$. Then, even with the extra rewards, the agent will suffer a total loss if she accepts every bet offer. She will lose $1 - 3\varepsilon$ if *E* is true and $1 - 2\varepsilon$ if *E* is false.

For the agent to conclude that she has no reason to abstain from the bets offered at *t*, it is enough if (i) she believes her actions at *t* won't influence the potential bet offer at *t'* (which follows if the bookie is known to be persistent), and (ii) she expects to deal with the offer at *t'* in the same way independently of what she might do at *t*. As (i) and (ii) imply that her present actions won't influence her opportunities and behavior in the future, she will conclude for each of the two bets offered at *t* that buying this bet is preferable to abstaining, as it improves her prospects by ε independently of what she is going to do at *t'*. If she is synchronically, but not diachronically unified, she will consider the two bets offered at *t* jointly instead of separately, but still conclude that buying both of them is preferable since it improves he prospects by 2ε in comparison

[10] In terms of the agent's *prior* probabilities (at *t*), the prospect of selling bet 3 at *t'* has a negative expected monetary value: $P(E)(C_{\text{bet 3}} - (P(A/E) \times S_{\text{bet 3}})) = 1/5(10 - (¾ \times 20)) = -1$.

[11] Actually, in that paper Skyrms does not explicitly discuss violations of the Principle of Reflection. Instead, he focuses on an analogous objection to a diachronic Dutch book against an agent whose updating strategy violates the rule of Conditioning. This means that for some potential evidence *E* and some proposition *A*, the agent's strategy requires her to update her probability for *A* upon learning *E* to $P_E(A)$, where $P_E(A) \neq P(A/E)$. Still, Skyrms's idea can be used just as well to deal with the objection at hand.

with buying none and by ε in comparison with buying just one, whatever she is going to do at t'.[12]

The key to this diagnosis is that, in the envisaged set-up, the agent cannot prevent the later offer of a bet by refusing the earlier bets. And if the offer will be made, it will be attractive in terms of the agent's probabilities at that future time. Under these circumstances accepting the current bet offers is preferable, since it diminishes the agent's expected loss. As Skyrms puts it:

> Why is it assumed [by Maher and Levi] that the cunning bettor will just go home if [the agent] refuses to bet today? […] Even though [the agent] will see it coming, she will prefer the sure loss […] because doing so looks strictly better to her than the alternative. (Skyrms 1993, pp. 323f)

And he concludes: "'Seeing it coming' does not help." (ibid., p. 326)

What would help, then? A salient feature of this case is the agent's disunification over time. The violator of Reflection is exploited because she decides on different bets at the times they are being offered, instead of making one decision on all the three bets together. If she did the latter, then – assuming she is logically and mathematically competent and prefers to be better off rather than worse off – she would certainly not choose to accept the whole bet package, since a simple calculation would show that refusing the three bets would be better for her whatever happens.[13]

3. Money Pumps against agents with cyclical preferences

Suppose an agent's preferences (represented by $\prec$) over alternative outcomes *x*, *y*, and *z* are cyclic:

$$x \prec y \prec z \prec x.$$

Let *x* be the *status quo* alternative. The agent is offered *y* in exchange for *x*, provided she pays $\varepsilon > 0$ for the exchange, where that payment is too small to reverse her preferences over outcomes. After this trade, she is offered to trade *y* for *z*, if she pays

---

[12] This reasoning appeals to (statewise) *dominance*. An action *A* is preferable to an action *B*, if *A* would lead to a better result than *B* in every state, i.e. whatever happens.

[13] Unification is not the only thing that would help in the present case. A disunified agent who violates Reflection could avoid exploitation if, contrary to what we have assumed, her earlier choices had an appropriate causal influence on her later behaviour. For example, she wouldn't be vulnerable to exploitation if her acceptance of the earlier bets would cause her to reject the bet offered at the later stage. But it's not safe to rely on causal connections between choices: They could just as well work to the agent's disadvantage: Acceptance of the earlier bets could cause the agent to accept the later bet as well. Unification in decision making is clearly superior to causation in this respect. (Similar remarks apply, *mutatis mutandis*, to agents with cyclic preferences that we are going to discuss in the next section.)

an additional ε. If she accepts, she is then offered to trade z for x, provided she again pays ε. After the three exchanges, the agent is back to where she started, minus 3ε. She has been used as a money pump. (Cf. Davidson, McKinsey and Suppes 1955, and Raiffa 1968).

For this pump to work, the extra payment of ε should not reverse the agent's preferences at any stage, at least up to 3ε. Thus, we need to assume that

$$x \prec y - \varepsilon \prec z - 2\varepsilon \prec x - 3\varepsilon.$$

The money-pump argument, as described above, invited an obvious objection: For the pump to work, the agent mustn't know she is being taken for a ride. Otherwise, if she knew that further trades are being kept in store for her, she would refuse to trade (cf. Schick 1986, and Schwartz 1986). The objection is that the condition of foresight, which should be satisfied in diachronic pragmatic arguments, is not satisfied in the money pump in its traditional version. A prudent agent with foresight would avoid to be pumped, because she would see what's coming. (If, after having read the preceding section, you already see that this objection can be dealt with, please be patient with with my discussion of this case. The two cases are similar, but they are not fully analogous.)

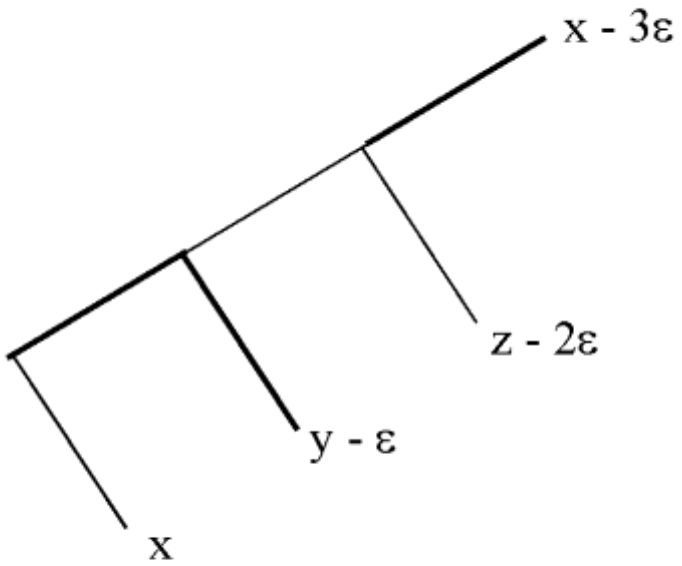
The idea of foresight coupled with prudence as a shield against exploitation can be made more precise in terms of *backward-induction* reasoning. When an agent confronts a sequential choice problem and has a robust trust in her future practical rationality, with the latter being interpreted as itself being a robust feature of the agent – a feature she would exhibit at any future moment of choice, independently of her past performance – she can solve the problem she faces by reasoning backwards, so to speak. She can first determine what move it would be rational for her to make at the last choice node at each branch of her decision tree, where it is clear what payoff her move would result in. Relying on her future robust rationality, she can predict she would make that move were she to reach the node in question. Taking her trust in her future rationality also to be robust, she expects to hold on to these predictions upon reaching the next-to-last choice node on each branch. This allows her to determine what move would be rational at each such penultimate node and thus, again relying on her future robust rationality, to predict her own behavior at that node. Continuing in this way, from the end-points of the tree to its beginning, such a *sophisticated chooser*

finds out what moves are rational at each choice node of the tree. To put it shortly: At each choice node, the backward-induction move is the one that would be optimal on the assumption that any move made at that node would be followed by the backward-induction moves at all the later choice nodes. (Note, though, that this simple formulation ignores the possibility of ties. With potential ties, the definition of backward induction is more complicated.)

Backward-induction reasoning is readily applicable to money-pump problems. As was argued by McClennen (1990, section 10.2), a sophisticated chooser – i.e., a disunified but farsighted agent, who puts her foresight to use in backward induction reasoning – will avoid being pumped. I argued for the same claim myself in Rabinowicz (1995). Since McClennen's original argument was slightly flawed, the presentation below follows my 1995 paper.

We consider the agent's sequential choice problem that consists of three trade offers in a row:

Figure 1: Money Pump



The forks in this decision tree are the agent's choice nodes. Going up means trading, going down is refusing to trade. The final outcomes are specified at the end-points of each branch in the tree. The bold lines represent backward-induction moves. At the third node, the agent's preferences dictate trading (i.e. going up), since she prefers $x - 3\varepsilon$ to $z - 2\varepsilon$. Given that she expects to trade at the third node if she were to reach that far, her choice at the second choice node should be to *refuse* to trade (i.e. to go down):

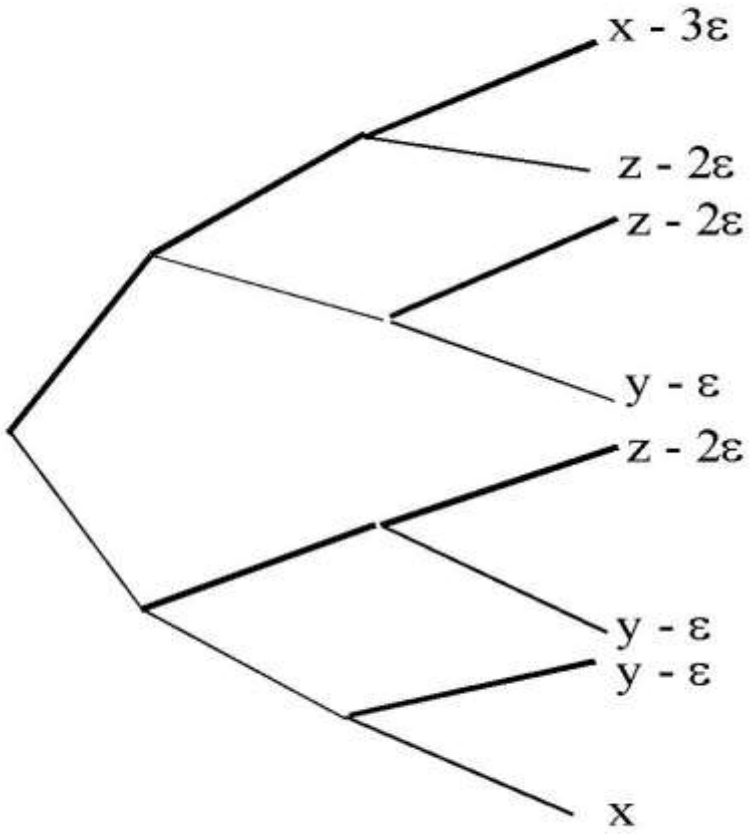
This gives her $y - \varepsilon$, which she prefers to $x - 3\varepsilon$. But then, given that she expects to refuse at the second node, her choice at the first node should be to *trade*, since she prefers $y - \varepsilon$ to x. Thus, the sophisticated chooser will make one exchange but then move no further. Even though her preferences are cyclical, she will not be pumped.

Are we then out of the woods? Is foresight coupled with prudence sufficient to stop the pump? Not quite. What follows is a description of a money pump that can be used against a sophisticated chooser (cf. Rabinowicz 2000).

In the money pumps discussed up to now, the series of trades terminates as soon as the agent refuses to make yet another exchange. No further trade offers are forthcoming. Suppose we change this feature of the decision problem and assume the exploiter to be *persistent*: If the agent refuses to trade, the exploiter comes back with the same trade offer at the next stage.[14] There are three stages at which offers are made. The decision tree for this new money pump looks as follows:

[14] Obviously, it is a variant of the same idea that was exploited by Skyrms (1993) in his treatment of diachronic Dutch books (cf. the preceding section).

Figure 2: Money Pump with Persistent Offers



As before, trades and refusals to trade are represented as upward and downward moves, respectively. If the agent each time refuses to trade, she ends up with *x*. If she trades just once, at whatever stage, she ends up with $y - \varepsilon$. If she trades twice, she receives $z - 2\varepsilon$. Finally, if she trades three times, she receives $x - 3\varepsilon$, i.e. gets back to where she has started, minus extra payments.

The bold lines again stand for the backward-induction moves. It is easy to see that the following holds:

(i) At each ultimate choice node, backward induction prescribes trading, as this gives the agent her preferred alternative and she knows that her choice is terminal: No further trade offers will be forthcoming.

(ii) Since she predicts she will trade at each ultimate node and since she expects to hold on to this prediction, she should also trade at each penultimate node, for the following reason: For the upper penultimate node, she predicts that trading at that point would eventually lead to $x - 3\varepsilon$ while refusal to $z - 2\varepsilon$, which she disprefers to $x -$

3ε. Analogously, for the lower penultimate choice node, she predicts that trading at that node would eventually lead to $z - 2\varepsilon$ while refusal to $y - \varepsilon$, which she disprefers to $z - 2\varepsilon$.

(iii) Given that she predicts she will trade at each subsequent node, she should trade at the first node as well. Trading at that node would eventually lead to $x - 3\varepsilon$, while refusal would lead to $z - 2\varepsilon$, which she disprefers to $x - 3\varepsilon$.

We conclude, then, that in this modified money pump, a sophisticated chooser with cyclic preferences will be pumped: She will trade each time, which will get her back to where she started, minus extra payments. The reason this pump works is obvious. The exploiter, being persistent in his offers, never lets the agent off the hook. Refusal to trade at an early stage does not terminate the pump: The trade offer will instead be repeated.

That backward induction implies repeated trading, if the exploiter is persistent, is a robust result, which can be generalized to pumps with an arbitrary number of stages (for the proof, see Rabinowicz 2000). Such pumps may be based on any number n of basic cycling alternatives, $x_1,\ldots, x_n$ (in our example, $n = 3$), and they may involve any number k of full rounds (in our example, $k = 1$). The only extra assumption we need to obtain this result is that the small payment required by each trade never reverses the agent's preference with regard to the basic cycling alternatives, independently of how many such payments she has already made.

There are obvious similarities between this set-up and the one considered by Skyrms. In both cases, the exploitation is made possible by the persistency of the exploiter. To be sure, this notion of persistency is cashed out in different ways in the two set-ups. In Skyrms's scheme, the exploiter is persistent in the sense that his later bet offers are not conditioned on the agent's acceptance of the earlier offers. In my scheme, the exploiter is persistent in the sense that he repeats the exchange offers that have been rejected by the agent. But the main idea is the same in both cases: Persistency is a feature that prevents the agent from getting off the hook by actions that would stop the ongoing exploitation scheme before it has run its full course.

A difference between the current set-up and the one suggested by Skyrms is that the Skyrmsian agent does not need to make use of backward induction. Accepting the earlier bets is advantageous for her whether or not she is going to accept the bet that will be offered at the later stage if E will turn to be the case. It is different in our

money pump. For example, trading at the first stage is advantageous for the agent because she is going to trade at both stages that follow. Were she instead to trade only once in what follows, it would be rational for her to abstain from trading at the first node.

As is well-known, the method of backward induction is quite controversial. The assumptions needed for its defense, as a general method of solving extensive-form games and dynamic decision problems, are very strong. Too strong, many would say. In particular, as mentioned above, the agent is assumed to to have a robust trust in her own future rationality and the future rationality of her protagonists. And she must expect to keep this trust under all counterfactual circumstances, given all possible evidence about past behavior. That is, she must expect to keep it even at the choice nodes that can only be reached by a sequence of irrational moves, on her part or on the part of other players. But, intuitively speaking, at such choice nodes one would expect her trust in the future rationality of the players to be undermined. (Note that these choices might be irrational not only intuitively, but also by the standards of backward induction itself.)

However, in Rabinowicz (1998), I have shown that a defense of backward induction for a limited class of (what I have called) *terminating* games and decision problems can be based on much less controversial assumptions (cf. also Aumann 1998 and Broome &Rabinowicz 1999 for further discussion). A game or a decision problem is terminating, if, at each of its choice nodes, backward induction prescribes a terminating move, i.e. a move that is not followed by any further moves. It can be shown that to obtain the backward-induction solution for games of this kind, there is no need for robusttrust inrationalit. It is enough to assume that each player expects to retain her original trust in rationality of the players as long as she lacks evidence to the contrary, i.e., as long as no irrational moves are made in the game.

Obviously, the Money Pump with Persistent Offers is not a terminating decision problem. However, in Rabinowicz (2001), I have shown that it is possible to set up a *terminating* decision problem – a "centipede for intransitive preferrers" – in which an agent with cyclic preferences will *forgo sure benefits* if she solves her problem using backward induction. (Forgoing sure benefits is just as bad, one might say, as accepting sure losses, which is the predicament of the exploited agent.)

The problem in question is just like the standard Money Pump from Figure 1, apart from two differences: (i) Each offer, if accepted, involves exchanging what one holds

for a *dis*preferred alternative: x for z, then z for y, and finally y for x. (ii) Instead of having to pay for an exchange, there is a small *reward* each time, with the proviso, however, that these small rewards do not suffice to reverse the agent' preferences. I.e., $x \succ z + \varepsilon \succ y + 2\varepsilon \succ x + 3\varepsilon$. As soon as an exchange offer is rejected by the agent, no further exchange opportunities are forthcoming. Thus, there is no persistency in offers. As can easily be seen, the backward-induction solution prescribes rejecting exchange offers at each choice node and thereby terminating the interaction. Thus, using backward-induction the agent goes down in the first move, thereby forgoing a sure benefit: If she instead made the three exchanges, she would get back to where she started while gaining 3ε on the way.[15]

Let us go back, however, to the main thread of our discussion. As in the set-ups in the two preceding sections, the agent with cyclic preferences is exploited in the Money Pump with Persistent Offers because her decision-making is disunified: She decides on each exchange separately, at the stage at which it is being offered, instead of making a single choice concerning all the three stages. Were she to make a single choice, then, we may safely assume, she would not choose to accept all the three exchanges: A simple calculation would show that refusing all of them would save her the extra costs and still result in the same outcome (x).

Since her preferences are cyclic, it is not determined by our description of the case what particular outcome the unified agent *would* choose in such a situation. But this cyclicity in her pairwise preferences does not imply that she would be unable to make a rational choice, when she considered all the alternatives together.

Here's how one can think of this problem. Let C be a choice function that picks out subsets from sets of alternatives in a given domain. Intuitively, for any alternative set S, C(S) consists of all the alternatives in S that the agent would view as choiceworthy if she were confronted with S as the set of alternatives to choose from. We allow that C(S) might be be empty for some non-empty sets S in the domain. Pairwise preference can be defined in terms of C: An alternative i is (strictly) preferred to an alternative j if and only if C($\{i, j\}$) = $\{i\}$. Analogously, indifference (equipreference) between i and j means that C($\{i, j\}$) = $\{i, j\}$. If C($\{i, j\}$) is empty, there is a gap in the agent's

[15] Yet another version of a money pump has recently been proposed by Dougherty (1913). His elegant explotation set-up only appeals to (statewise) dominance and does not rely on backward induction at all. In this respect it is like Skyrms's construction. However, it only works against agents who have cyclic preferences over packages consisting of several components and in addition requires that their preferences over components are weakly separable. Thus, it is limited in generality.

preference ordering as far as the comparison between *i* and *j* is concerned. Note that, on this approach, the notion of a choice function C is definitionally prior to the notion of preference. Thus, we do *not* assume that C(*S*) is definable as, say, the set of alternatives in *S* that are 'optimal', i.e. preferred to or equipreferred with every alternative in *S*, or, more cautiously but still contentiously, as the set of alternatives in *S* that are 'maximal', i.e. not dispreferred to any alternative in *S*. Indeed, we do not even require that C(*S*) is 'closed upwards', i.e., that an alternative in *S* that is preferred to some alternative in C(*S*) must itself belong to C(*S*). Admittedly, this may be viewed as an extremely permissive approach to choiceworthiness, but perhaps not excessively so.

Consider a set *S* of cycling alternatives. For any alternative *j* in *S*, *S* contains some *i* such that *i* is preferred to *j*. However, this is compatible with C(*S*) being non-empty. In such a case, the cycle in *S* can be said to be *benign*. If, on the other hand, C(*S*) is empty, the cycle is *vicious*: There is then no room for a rational choice from the set of cycling alternatives.[16]

In our example of the money pump, it is possible that the cycle indeed might be benign. But $x - 3\varepsilon$ will certainly not be among the choiceworthy alternatives, i.e., it will not belong to *C*(*S*), given that it is worse than *x* in one respect (-3ε) and is just like *x* in every other respect. It might still be the case that *C*(*S*) is non-empty; for example, *x* itself might belong to *C*(*S*). But if *C*(*S*) is non-empty, it cannot be closed upwards. If it were, then *C(S*) would have to contain all the cycling alternatives, including $x - 3\varepsilon$, as soon as it contains one of them.

The distinction between the two types of cyclicity – the benign and the vicious one – is important when it comes to the discussion of the rationality of cyclical preferences. Benign cyclicity allows for a rational choice from the cycle as a whole. Thus, a unified agent with benign cyclic preferences might rationally choose one of the choiceworthy alternatives in the cycle and head on to it. Since $x - 3\varepsilon$ is not choiceworthy and therefore will not be her chosen alternative, she will not be pumped. Things are different for a disunified agent, even if the latter is sophisticated enough to use backward induction in her practical reasoning. Such an agent will *still* be subject to a money pump, quite independently of whether the cycle in her

[16] On this issue, see Rabinowicz (2000). The present treatment slightly differs from the one in that paper.

preferences is benign or vicious. And even if it is vicious, she will still act rationally while getting money-pumped. The reason is that at each choice node she is faced with a choice between two final outcomes only. She never has to choose from the whole cycling set of the three final outcomes.

I mention this possibility of benign cycles, because Levi (2006, pp. 375f) does not take it into consideration. There is no need, according to him, to resort to money pumps in criticizing cyclic preferences. What makes such preferences unacceptable in his view is the agent's precarious predicament when it comes to a choice from the whole set of cycling alternatives:

> Cyclic preferences are irrational precisely because *X* [who has such preferences] cannot choose rationally in some decision problems. Were *X* confronted with a three way choice between *x, y* and *z, X* could not follow the policy of choosing an option that is […] optimal according to some permissible ranking and, indeed, could not follow the slightly different policy of choosing an option that is maximal in the sense that no option is strictly preferred to it. I am convinced by this argument that cycles should be avoided. Rabinowicz's argument [= my money pump with persistent offers] seems far less compelling. [17]

It should be clear why I don't think that this criticism of cyclicity is convincing. Even though every option in the cycling set is dispreferred to some of its competitors, it may be that the cycle in question is benign, i.e. that some of the options in the cycle are choiceworthy despite being neither optimal nor maximal. In that case, a rational choice from the set of cycling options is possible, contrary to what Levi suggests.

4. Levi's criticism of diachronic pragmatic arguments

In "Money Pumps and Diachronic Dutch Books" (2002), Levi considers my money pump with persistent offers and Skyrms's version of a diachronic Dutch book with a persistent exploiter. He argues that there is a decisive difference between these diachronic exploitation schemes and *synchronic* exploitation set-ups, such as classic synchronic Dutch books. The difference has to do with the range of options that are available to the agent. The actions of an agent who is being exploited in a synchronic set-up are, when taken together, (statewise) *dominated by an option that stands at her*

[17] Cf. also Levi (2002), p. 242:

> Let Z have strict categorical preferences for A, B, and C that yield a cycle. How should Z choose when all and only these three are options available to him? Maximizers of value will refuse to choose any option dispreferred to all other options […] By this consideration, none of the options may be recommended. But decision-makers who evaluate their options so that no option available to them is admissible are synchronically incoherent. It is a cardinal condition of rational choice that the set of admissible options be nonempty. One should avoid cycles in categorical preference to avoid violating this cardinal requirement.

disposal. The agent accepts each bet in the book even though she could have chosen an option – to reject all the bets on offer – that would have led to a better outcome under all possible circumstances – in every possible state of the world. In a diachronic set-up, things are different. Thus, consider the agent *X* at the initial choice node. "*X* has no control then over what [she] will choose later. He can only predict what he will do." (Levi 2002, p. 239) If he is exposed to the money pump with persistent offers and ends up making the three trades, "*X* is not choosing [at any point] an option dominated by another *available as an option* to him" (ibid., p. 241, Levi's emphasis). In particular, at the initial choice node, refusing to trade at any of the three stages is not an option that stands at the agent's disposal. Because of this absence of an available option that dominates the course of action actually taken, the agent cannot be charged with irrationality.

To be sure, Levi writes, a money pump with persistent offers shows that an agent with cyclic preferences can be 'taxed' for having preferences of this kind. The extra costs she incurs may be seen as tax payments. But vulnerability to taxation is not irrationality. Levi concludes:

> It may be argued that if X did not have the cyclical preferences, he could not be taxed in the manner just sketched. Avoiding cyclical preferences to avoid taxation is not avoiding a dominated option. It is adapting one's preferences to circumstances as in the case of sour grapes.
>
> Is adjusting preferences so that one may not consider oneself a victim of taxation a good idea? I doubt whether a general all-purpose answer can be given to this question. It is certainly not a requirement of minimal rationality. Why should we mandate sour grapes? […]
>
> Money Pump arguments were designed initially to show that individuals who violate certain canons of rationality will end up choosing options that are dominated by other options available to them *just like synchronic arguments do*. Showing that violating these canons is one way, that in the face of other assumptions, makes one vulnerable to taxation, is no substitute. Those who use money pump arguments to defend acyclicity of preference have failed to show that decision-makers who violate acyclicity are driven to choose dominated options. (ibid., pp. 241f)

Levi's diagnosis of Skyrms's version of the diachronic Dutch book is exactly analogous. The agent lacks control over her future choices; she can only predict what she will do. Consequently, she cannot at *t* decide to refuse all the bets offers, both the ones made at *t* and the ones she might receive at *t'*. This means that she cannot be accused of acting in a way that is dominated by some option that stands at her disposal.

> Diachronic dutch books purport to show that the decision maker X at the initial node will be driven by considerations of rationality to choose an option inferior to some other option available to him no matter what is the case consonant with X's initial state of full belief.
>
> According to Skyrms's scenario, *X* is worse off, no matter how *X* chooses, than *X* was in the initial status quo. If *X* has the option of remaining in the status quo position, *X* should do so [rather than act as she does]. But by hypothesis *X* does not have this option. […] Buying [the bet

> on *A* conditional on *E*] at the initial stage is not dominated by refusing to buy it at that stage. Since these are the only two options, where is the beef? (ibid., p. 247)

5. My response

Indeed, where is the beef? Levi is quite right that, both in my money pump and in Skyrms's diachronic set-up, it is assumed that the agent at the initial stage cannot control what she will do in the future. As Skyrms puts it: "Deciding not to bet ever is not an option." (Skyrms 1993, p. 323) Consequently, the agent's course of action is not dominated by any of the options that stand at her disposal. It is only dominated by a certain *sequence* of options, which are available to the agent at different times. But the sequence as a whole is not an option for the agent, at any time.[18]

However, to deal with this issue, we can simply modify the diachronic set-ups so as to put the two kinds of arguments, the synchronic and the diachronic kind, on an equal footing. Let us assume, therefore, that the agent at the initial stage *can* decide on the whole temporal sequence of her actions, but, as a matter of fact, she never does and instead decides on different offers at the times when they are made. However, if she did view her decision-problem in a unified way, which she could do, her prior planning decision concerning the whole action sequence would determine her subsequent behavior.

In this way, the synchronic and the diachronic exploitation set-ups become analogous. In the synchronic case, the agent is also assumed to engage in a disunified decision making: She makes decisions on each bet separately. (Otherwise, as we have seen, no exploitation would take place.) But, if she viewed the situation in a unified manner, she would then make a single choice as to which bets to accept and which to reject. It is in this sense that she has at her disposal the option of declining all the bets, which dominates her actual behavior. This option is available to her, since it would figure in her deliberation as one of the alternatives if she were unified (which she is not) and nothing hinders her from viewing the decision problem in such a unified fashion. In this respect, then, the synchronic set-up is similar to the diachronic one,

[18] That the absence of a feasible dominating option in diachronic exploitation set-ups makes the latter inadequate as means to establish the irrationality of constraint violations is also emphasized by Teddy Seidenfeld in his writngs (cf. Seidenfeld 1988, pp. 280f). Steele (2010, p. 274) summarizes Seidenfeld's position as follows: "The dominating strategy against which we measure the agent's sure loss in each of the scenarios modelled above is not, in fact, a dynamically feasible option – the agent predicts that they would not make the requisite series of choices at the given choice nodes. The idea is that such infeasible options should not enter into any analysis of the decision problem and its solutions."

after we have modified the latter to make unified decision-making possible in this case as well. (For this suggestion, see Rabinowicz (2006).)

One might point out that there still is this difference between the synchronic and the diachronic case: In the diachronic case, when I separately consider each offer, I need to predict my future choices regarding later offers in order to determine the final outcome of my current choice. In the synchronic case, however, when considering an offer of a bet, I do not make any predictions about the decisions I take regarding other bet offers in the package. As long as each of the other offers still is under deliberation, I cannot – it seems – relate to them in a predictive mode. At least on one interpretation of Levi's thesis that deliberation crowds out prediction, this is, I guess, what he would want to say.[19] But then disunification in the synchronic case involves more than just separate decisions on each bet offer. It would also seem to involve some form of abstraction from the context: While considering whether to accept a given bet, the agent disregards the other bets that are on offer and her decisions on those bets.[20]

This difference between disunification in the diachronic and in the synchronic case, is important. But, if anything, it makes synchronic arguments *less* compelling than the diachronic ones. The kind of disunification in decision-making that is required for the former arguments to work involves putting artificial blinkers on the deliberating agent. It is thus a more radical and for that reason more problematic form of disunity than the one needed for the latter arguments. [21] And, anyway, this difference does not

---

[19] For his exposition and defence of that thesis, see Levi (1989), (1991), (1997). For a critical discussion, see Joyce (2002) and Rabinowicz (2002).

[20] In the diachronic case, an analogue of such radical disunification can be found in *myopic* choice: A myopic agent chooses between current options without taking into account her future action opportunities; she approaches her current decision as though it were the only decision she is ever going to make. As we have seen, myopia is not a pre-requisite for exploitation in the diachronic cases. Foresight alone does not suffice as a safeguard.

[21] An example of how this difference can play a role is provided by a recent discussion of the question whether subjective probabilities must be sharp. Elga (2010) constructed a kind of 'reverse' diachronic Dutch book against an agent with imprecise subjective probabilities. The book consisted of two bets, offered in sequence, one on Heads and the other one on Tails in the same throw of the coin, where the agent's probabilities for these events were supposed to be imprecise. Both bets had the same price and stake (price: 10, stake: 25), chosen in such a way that the agent who accepted both bets would benefit whatever happens (she would pay 10+10 for the two bets, but would win 25 whatever happened). A Dutch book of this 'reverse' kind is successful if the agent refuses each of the bets in the book, thereby forgoing a sure benefit. (Cf. the discussion of the centipede for intransitive preferrers in section 3 above.) In particular, Elga suggested that his Dutch book would be successful if the agent with imprecise probability assignments followed the rule of maximizating minimal expected utility (MmEU). If the agent's probability for Heads is given by an interval from, say, .2 to .8 (which induces the corresponding interval for Tails), the minimal expected utility for each of the two bets is negative ($0.2\times25 - 10 = -7{,}5$).

affect the issue of the availability of a dominating option. In both set-ups, the dominating option might be available without being an alternative that the agent considers in her (disunified) deliberation. If this is the case, then Levi no longer has grounds for his claim that synchronic pragmatic arguments have a bite that the diachronic arguments lack.

In the synchronic set-up, the presence of the option to refuse all the bets in the package does not, on the received view, make it irrational for the agent to decide to accept any particular bet, when she considers whether to accept it or not. The agent views her decision problem in a disunified fashion, in which the option of the wholesale refusal does not figure as one of the alternatives. The same should therefore apply to the diachronic case. In the diachronic set-up, the mere presence of the option to refuse all the offers, the current and the future ones, does not make it irrational for the agent to accept any particular offer in the exploitation sequence, when she considers whether to accept that offer or to refuse it. For, again, she views her decision problem in a disunified fashion, which means that the option of the wholesale refusal does not figure as one of the alternatives she considers.

In his more recent comments, Levi (2006) addresses this point. Commenting on the above suggestion and using my money pump with persistent offers as an example, he argues that the disunified form of practical deliberation *as such* must be irrational if it is not inescapable:

As has been shown by Sahlin and Weirich (2013), Elga was mistaken in his diagnosis. The 'reverse' diachronic Dutch book doesn't work against the (disunified) MmEU agent, if the latter makes use of backward induction. Then, as is easy to see, she is going to accept each of the bets in the book. More precisely, if she won't accept the first bet, she is not going to accept the second one either, given that its minimal expected utility is negative. But if she will accept the first bet, she is going to accept the second one. The reason is that accepting the second bet in addition to the first one will guarantee her a sure total gain. Relying on her future rationality, she therefore has a good reason to accept the first bet as well. Elga (2012) recognized the validity of this criticism.

However, what Elga didn't point out is that his 'reverse' book would *still* work against a disunified MmEU-agent, if it were set up *synchronically* rather than diachronically. Thus, it would still work if all the bets in the book were offered to the agent at the same time rather than sequentially. As each bet is unattractive *on its own* by MmEU lights, it will be rejected in this synchronic set-up unless the agent decides on both offers together rather than on each of them separately. (In fact, in their paper, Sahlin and Weirich implicitly assume that a rational agent must be synchronically unified and therefore dismiss the synchronic construction of Elga's argument with the following remark: "An agent deciding simultaneously about A and about B has the option of accepting both gambles and the option of rejecting both gambles because a combination of compatible options at a time is also an option at the time. In a synchronic version of Elga's problem, MmEU prohibits rejecting both gambles because the minimum expected utility of accepting both gambles is greater than the minimum expected utility of rejecting both gambles.")

> A possible way to understand Rabinowicz's suggestion is that *X* has control *at the initial node* over which of the eight paths *X* will choose. [Here, "the eight paths" refers to the eight branches in the decision tree for the money pump in question.] But *X* deliberates in a "disunified" way so that at each node he deliberates between the "sell-don't sell" options available then.
>
> If *X* refuses to consider all the options that are available to *X* according to *X*'s beliefs and goals, *X*'s deliberation is irrational. Indeed, this is so whether or not the options that are not considered dominate the one chosen from the options that are. Such disunity is to be avoided. This is so whether *X* is offered a set of gambles at the same time or is offered a sequence of options where *X* regards *X* to be in control of the path *X* will take. (Levi 2006, p. 376)

Levi's point, then, is that any deliberation that ignores some of the options available to the agent is *ipso facto* irrational. Since disunified deliberation has this feature (as it ignores the 'wholesale' options), it is always irrational. It is irrational whether or not the ignored options dominate the chosen ones and whether or not the decision set-up is diachronic or synchronic.

If Levi is right, pragmatic arguments for various constraints on the agent's beliefs and desires would *all* seem ill-conceived insofar as they involve exploitation set-ups that consist of collections of opportunities. We have seen that such arguments presuppose not only that the agent violates the relevant constraints on her beliefs and preferences but also that she makes her decisions in a disunified fashion. But then, if disunification *itself* is irrational since it makes the agent ignore some of the available options, the arguments lose their bite: They provide no reasons for upholding the relevant constraints. Instead, all blame falls on disunification as such.[22] Note also that, if Levi is right in what he now suggests, then synchronic and diachronic arguments would again be put on equal footing. They would all be equally worthless.

But is Levi right? Is it always irrational to ignore some of the options that are available for choice? I very much doubt it. It is one thing to require that the considered action alternatives should be *jointly exhaustive* in the sense that the agent is bound to perform (at least) one of them in every possible development. It is quite another thing to demand that the agent should consider every available alternative. When I deliberate in a disunified fashion whether or not to accept a certain opportunity, the alternatives I consider are jointly exhaustive, despite the fact that I ignore the more complex options, which concern not only this opportunity but also other opportunities as well. This, I imagine, is how things are with all of us, most of the time. In practically every choice situation, the alternatives that figure in our

[22] Note that on this proposal unification in decision making does not come out unscathed either. A unified agent considers wholesale options, which means that she disregards more limited kinds of opportunities (such as accepting or rejecting a particular bet offer), despite the fact that such options also are available. This cannot be rational either on Levi's (2006) view.

deliberation admit of versions and extensions that we do not reflect upon. However detailed our deliberation might be, there always is room for more detail and elaboration. Thus, Levi demands too much.

We can, to be sure, require that the set of alternatives the agent considers is not too meagre – that it obeys *some* richness conditions. One such condition has already been mentioned: The set in question should be jointly exhaustive. Another condition that seems reasonable is (weak) closure under dominance: if the agent considers an action *A*, and there are available actions that dominate *A*, then at least one of such actions should belong to the set of alternatives the agent considers. (A strong closure under dominance would require the agent to consider all the actions that dominate *A*, but this seems much too strong: there may be indefinitely many such actions.) Now, one might think that closure under dominance is violated whenever disunification leads to exploitation. This, however, would be a mistake. It is true that the disunified agent in our examples does not consider the option to refuse every offer, even though this wholesale refusal dominates the option to accept all the offers. However, the latter option is not among the ones he considers either. Being disunified, his options are piecemeal rather than wholesale: his deliberation is always restricted to the question whether to accept a particular offer or to reject it. Since the option to refuse all offers does not dominate the option to accept a particular offer, closure under dominance is not violated in such cases.

6. Summing up

Let me sum up the main claims of this paper. Pragmatic arguments cannot establish the inherent rationality of constraints on the agent's state of mind. Instead, the kind of rationality they target is purely instrumental: Their proper role is to identify constraints that the agent has reason to comply with in order to safely engage in disunified decision-making. In Levi's apt terminology, it is a matter of 'tax avoidance': Pragmatic arguments identify constraints one needs to satisfy to be shielded against the tax imposed on disunification. I share Levi's view that avoiding tax at all costs is unreasonable, especially if it is a matter of constraints that do not seem to be inherently compelling. (The Principle of Reflection is a case in point.) I do *not* share his more recent view that disunification as such is always irrational just because it involves ignoring some available options. However, to the extent that (i)

synchronic unification is much easier and less costly to achieve than its diachronic counterpart, and also because (ii) synchronic disunification requires resorting to artificial blinkers in deliberation (it requires disregarding one's decisions concerning other bet offers when considering a given bet), *diachronic pragmatic arguments provide us with stronger instrumental reasons for compliance*. Unlike synchronic disunification, which appears to be a purely theoretical construct, diachronic disunification is a fact of life. This gives special weight to diachronic pragmatic arguments which identify constraints that safeguard a diachronically disunified mind. There is some irony in this conclusion. If I am right, the classic and most influential pragmatic arguments - synchronic Dutch books - are considerably less compelling than their younger diachronic cousins.

References

Aumann, R., 1995, "Backward Induction and Common Knowledge of Rationality", *Games and Economic Behavior* 18: 6-19.

Bar-Hillel, M., and Margalit, A., 1988, "How Vicious are Cycles of Intransitive Choice?", *Theory and Decision* 24: 119-145.

Blamey, J., 2011, *When Do I get My Money? - A Probabilistic Theory of Knowledge*, Ph.D. thesis, Det. Of Philosophy, King's College London.

Bratman, M., 1987, *Intention, Plans, and Practical Reason*, Cambridge, Harvard University Press.

Bratman, M., 2007, *Structures of Agency: Essays*, New York, Oxford University Press.

Bratman, M. 2012, "Time, Rationality, and Self-Governance", *Philosophical Issues* 22: 73-88.

Broome, J., and Rabinowicz, W., 1999, "Backwards Induction in the Centipede", *Analysis* 59: 237-242.

Chignell, A., 2007, "Belief in Kant", *The Philosophical Review* 116: 323-360.

Christensen, D., 1991, "Clever Bookies and Coherent Beliefs", *The Philosophical Review* 50: 229-247.

Davidson, D., McKinsey, J.C.C., Suppes, P., 1955, "Outlines of a Formal Theory of Value, Part I", *Philosophy of Science* 22: 140-160.

Dougherty, T., 2013, "A Deluxe Money Pump", *Thought: A Journal of Philosophy*, article first published online, http://onlinelibrary.wiley.com/doi/10.1002/tht3.91/pdf

Elga, A., 2010, “Subjective Probabilities Should Be Sharp”, *Philosophers’ Imprint*, http://www.philosophersimprint.org/010005/

Elga, A., 2012, “Errata for *Subjective Probabilities Should Be Sharp*”, manuscript, http://www.princeton.edu/~adame/papers/sharp/sharp-errata.pdf

Eriksson, L., and Rabinowicz W., 2012, “The Interference Problem for the Betting Interpretation of Degrees of Belief”, *Synthese*, published on line.

de Finetti, B., 1990, *Theory of Probability*, Vol. I (first edition 1974), New York, Wiley.

Joyce, J. M., 2002, “Levi on Causal decision Theory and the Possibility of Predicting One’s Own Actions”, *Philosophical Studies* 110: 69-102.

Kant, Immanuel, 1904 [1787], *Kritik der Reinen Vernunft*, Akademieausgabe, vol. III http://www.korpora.org/kant/aa03/

Levi, I., 1988, “The Demons of Decision”, *The Monist* 70: 193-211.

Levi, I., 1989, “Rationality, Prediction, and Autonomous Choice”, *Canadian Journal of Philosophy* suppl.vol. 19: 339-63.

Levi, I., 1991, “Consequentialism and Sequential Choice”, in M. Bacharach and S. Hurley (eds.), *Foundations of Decision Theory*, Oxford, Blackwell: 92-122.

Levi, I., 1997, *The Covenant of Reason*, Cambridge, Cambridge University Press.

Levi, I., 2002, “Money Pumps and Diachronic Dutch Books”, *Philosophy of Science*, suppl. to vol. 69: s235-s247.

Levi, I., 2006, “How Adaptive Should Our Rational Preferences Be? Reply to Rabinowicz”, in Erik J. Olsson (ed.), *Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi*, Cambridge, Cambridge University Press: 375-78.

Maher, P., 1992, “Diachronic Rationality”, *Philosophy of Science* 59: 120-141.

McClennen, E., 1990, *Rationality and Dynamic Choice*, Cambridge, Cambridge University Press.

Moore, G. E., 1903, *Principia Ethica*, Cambridge, Cambridge University Press.

Rabinowicz, W., 1995, “To Have One’s Cake and Eat It, Too: Sequential Choice and Expected Utility Violations”, *The Journal of Philosophy* 92: 586-620.

Rabinowicz, W., 1998, “Grappling with the centipede”, *Economics and Philosophy* 14:94-125.

Rabinowicz, W., 2000, "Money Pump with Foresight", in Michael J. Almeida (ed.), *Imperceptible Harms and Benefits*, Dordrecht, Kluwer: 123-154.

Rabinowicz, W., 2001, ”A Centipede for Intransitive Preferrers”, *Studia Logica* 67: 167-178.

Rabinowicz, W., 2002, “Does Practical Deliberation Crowd Out Self-Prediction?”, *Erkenntnis* 57: 91-122.

Rabinowicz, W., 2006, “Levi on Money Pumps and Diachronic Dutch-Book Arguments”, in Erik J. Olsson (ed.), *Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi*, Cambridge, Cambridge University Press: 289-312.

Rabinowicz, W., 2008, “Pragmatic Arguments for Rationality Constraints”, in M.-C. Galavotti, R. Scazzieri and P. Suppes (eds), *Reasoning, Rationality and Probability*, Stanford, CSLI Publications/The University of Chicago Press: 139-163.

Rabinowicz, W., 2012, “Does Irrationality Make One Vulnerable to Exploitation?”, in V. Lektorsky and A. Guseynov (eds.), *Rationality and Its Limits*, Proceedings of the 2011 Moscow meeting of Institut International de Philosophie, publ. by Institute of Philosophy of the Russian Academy of Sciences: Moscow 2012, 67-89.

Raiffa, H., 1968, *Decision Analysis: Introductory Lectures on Choice under Uncertainty*. Reading, Mass., Addison-Wesley.

Ramsey, F. P., 1990 (1926), “Truth and Probability”, in D. H. Mellor, ed. *F. P. Ramsey: Philosophical Papers*, Cambridge, Cambridge University Press.

Sahlin N-E. and Weirich P., 2013, “Unsharp Sharpness”, *Theoria*, article first published online, DOI:10.1111/theo.12025

Seidenfeld, T., 1988, “Decision Theory Without ‘Independence’ or Without ‘Ordering’”, *Economics and Philosophy* 4, 267-290.

Schick, F., 1989, “Dutch Bookies and Money Pumps” *The Journal of Philosophy* 83: 112-119.

Schwartz, T., 1986, *The Logic of Collective Choice*, New York, Columbia University Press.

Skyrms, B., 1980, “Higher Order Degrees of Belief”, in D. H. Mellor, *Prospects for Pragmatism*, Cambridge, Cambridge University Press: 109-138.

Skyrms, B., 1993, “A Mistake in Dynamic Coherence Arguments?”, *Philosophy of Science* 60: 320-328.

Steele, K.S., 2010, “What Are the Minimal Requirements of Rational Choice? Arguments for Sequential Decision Setting”, *Theory and Decision* 68: 463-487.

Teller, P., 1973, “Conditionalization and Observation”, *Synthese* 26: 218-238.

Van Fraassen, B., 1984, “Belief and the Will”, *The Journal of Philosophy* 81: 235-256.