ROYAL INSTITUTE
OF TECHNOLOGY

PREFERENCE AND CHOICE

JOHAN E. GUSTAFSSON

Doctoral Thesis
Stockholm, Sweden 2011

Abstract Gustafsson, Johan E. 2011. Preference and Choice. *Theses in Philosophy from the Royal Institute of Technology 38*. 122 + x pp. Stockholm. ISBN 978-91-7415-951-6.

This thesis consists of five essays on decision theory and an introduction.

Essay I defends ratificationism from a recent attack by Andy Egan. Egan argues that neither evidential nor causal decision theory gives the intuitively right recommendation in the cases *The Smoking Lesion*, *The Psychopath Button*, and *The Three-Option Smoking Lesion*. Furthermore, Egan argues that we cannot avoid these problems by any kind of ratificationism. This essay develops a new version of ratificationism that yields the intuitively right recommendations. Thus, the new proposal has advantages over evidential and casual decision theory and standard ratificationist evidential decision theory.

Essay II develops a new version of the money-pump argument for the claim that rational preferences are transitive. The standard money pump only exploits agents with cyclic strict preferences. In order to pump agents who violate transitivity but without a cycle of strict preferences, one needs to somehow induce such a cycle. Methods for inducing cycles of strict preferences from non-cyclic violations of transitivity have been proposed in the literature, based either on offering the agent small monetary transaction premiums or on multi-dimensional preferences. This essay argues that previous proposals have been flawed and presents a new approach based on the dominance principle.

Essay III examines the small-improvement argument. This argument is usually considered the most powerful argument against completeness, namely, the view that for any two alternatives an agent is rationally required either to prefer one of the alternatives to the other or to be indifferent between them. The essay argues that while there might be reasons to believe each of the premises in the standard version of the small-improvement argument, there is a conflict between these reasons. As a result, the reasons do not provide support for believing the conjunction of the premises. Without support for the conjunction of the premises, the standard version of the small-improvement argument against completeness fails.

Essay IV models preference relations. In order to account for non-traditional preference relations the essay develops a new, richer framework for preference relations. This new framework provides characterizations of non-traditional preference relations, such as incommensurateness and instability, that may hold when neither preference nor indifference do. The new framework models relations with swaps, which are conceived of as transfers from one alternative state to another. The traditional framework analyses dyadic preference relations in terms of a hypothetical choice between the two compared alternatives. The swap framework extends this approach by analysing dyadic preference relations in terms of two hypothetical choices: the choice between keeping the first of the compared alternatives or swapping it for the second; and the choice between keeping the second alternative or swapping it for the first.

Essay V develops a new measure of freedom of choice based on the proposal that a set offers more freedom of choice than another if, and only if, the expected degree of dissimilarity between a random alternative from the set of possible alternatives and the most similar offered alternative in the set is smaller. Furthermore, a version of this measure is developed that is able to take into account the values of the possible options.

Keywords: preference relations; rationality constraints; transitivity; completeness; incommensurability; parity; money pumps; ratifiability; freedom of choice.

Johan E. Gustafsson, Division of Philosophy, Department of Philosophy and the History of Technology, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden

Typeset with LaTeX and Perl by the author (except essays I, II, III, and V). Written in Vim.

ISSN 1650-8831

ISBN 978-91-7415-951-6

This doctoral thesis consists of the following introduction and the essays:

I Gustafsson, Johan E.: 'A Note in Defence of Ratificationism', forthcoming in *Erkenntnis*.

II Gustafsson, Johan E.: 2010, 'A Money-Pump for Acyclic Intransitive Preferences', *Dialectica* **64**(2):251–257.

III Gustafsson, Johan E. & Espinoza, Nicolas: 2010, 'Conflicting Reasons in the Small-Improvement Argument', *The Philosophical Quarterly* **60**(241):754–763.

IV Gustafsson, Johan E.: 'An Extended Framework for Preference Relations', forthcoming in *Economics and Philosophy*.

V Gustafsson, Johan E.: 2010, 'Freedom of Choice and Expected Compromise', *Social Choice and Welfare* **25**(1):65–79.

CONTENTS

ACKNOWLEDGEMENTS

A large number of people have helped me during the course of this work. First and foremost, I am grateful to my supervisors, Sven Ove Hansson, Martin Peterson, and John Cantwell for their supererogatory assistance. Martin has also co-authored an article with me on a topic outside the scope of this thesis. I have moreover benefited from discussions with Nicolas Espinoza and with Karin Enflo who introduced me to the problem of measuring freedom of choice. Nicolas is also the co-author one of the essays in this thesis. I also wish to thank David Alm, Frank Arntzenius, Erik Carlson, Wlodek Rabinowicz, Tor Sandqvist, Fredrik Johansson Viklund, and Niklas Olsson-Yaouzis for their comments on earlier versions of some of the essays. Jesper Jerkert has been a constant source of opinions on the punctuation and spelling of my essays.

PREFACE

The aim of the introduction is not just to provide background for the essays. The essays were written to be read on their own and should, assuming that I have done my job, not need any introduction. Rather, the introduction is an attempt to offer some further thoughts on the topics of the essays. This has been an opportunity to develop some of my ideas a bit without the restrictions of a journal paper. I answer some objections that have surfaced since the publication of the essays. Furthermore, the results of some of the essays are used to form new combined arguments. Straightforward summaries of the essays follow after the introduction.

INTRODUCTION

Decision theory is the theory of rational decisions. Decision theory is hence concerned with what people rationally ought to decide rather than what people actually decide. More specifically, it is the theory of what decisions one is rationally required to make given one's preferences and beliefs. Thus, standpoints in decision theory do not imply any substantial rationality requirements about what particular things one should prefer or believe. That said, it is of major importance to decision theory what structure our beliefs and preferences are rationally required to have—classical decision theory implies a number of formal requirements on what combinations of preferences or beliefs one is rationally permitted to have.

The first two sections of this introduction cover some influential decision theories. None of these yields what seems to be the intuitively right recommendations in a series of decision problems that have been the focal point of much of the recent decision theoretical literature. In response to this I develop my own decision theory that gives the intuitively right recommendations.

Sections 3–7 examine the assumptions made by classical decision theories on the structures of the decision maker's preferences. A key issue in the foundations of decision theory is what preference relations are possible and what combinations of preferences are rationally permissible. This part of the thesis will examine arguments for and against standard requirements like transitivity and completeness. Furthermore, it will examine the arguments for non-traditional preference and value relations and different frameworks for making these conceptually possible. Finally, I will present my argument and framework for some non-traditional preference and value relations.

In addition to rational constraints there is a further, less easily violated, restriction on our choices. Your choices are always restricted by the range of alternatives available to you. Some of the time the limitations in what options are available may force you to make a less than ideal choice. Different sets of alternatives offer different amounts of freedom of choice. The question of how to evaluate the freedom of choice offered by a set of options is the topic of Section 8, which defends a new proposal. I argue that there is a connection between the amount of freedom of choice a set offers and how well it is expected to satisfy an agent with a certain kind of unknown preferences.

1. Newcomb problems

Some of the most discussed decision problems are the so called Newcomb problems. These problems have motivated some of the most important developments in decision theory. The first Newcomb problem was conceived by William Newcomb in 1960 while pondering the similarly structured Prisoners' Dilemma.[1] It was first published in a paper by Robert Nozick.[2] He presents the problem as follows:

> Suppose a being in whose power to predict your choices you have enormous confidence. [...] There are two boxes, (B1) and (B2). (B1) contains $ 1000. (B2) contains either $ 1000 000 ($ M), or nothing. What the content of (B2) depends upon will be described in a moment.
>
> (B1) {$ 1000} (B2) { $ M / or / $ 0 }
>
> You have a choice between two actions:
>
> (1) taking what is in both boxes
> (2) taking only what is in the second box.
>
> Furthermore, and you know this, the being knows that you know this, and so on:
>
> (I) If the being predicts you will take what is in both boxes, he does not put the $ M in the second box.
> (II) If the being predicts you will take only what is in the second box, he does put the $ M in the second box.
>
> The situation is as follows. First the being makes its prediction. Then it puts the $ M in the second box, or does not, depending upon what it has predicted. Then you make your choice. What do you do?[3]

In a recent survey by David Bourget and David Chalmers, professional philosophers were asked 'Newcomb's problem: one box or two boxes?' The results turned out as follows: [4]

[1] Gardner (1986, p. 156). For a discussion of the similarities between Newcomb Problems and Prisoners' Dilemma see Lewis (1979).

[2] Nozick (1969). However, Nozick wrote about the problem already in his dissertation Nozick (1963, p. 223), which was not published until 1990.

[3] Nozick (1969, pp. 114–115).

[4] Bourget and Chalmers (2009).

	Decision theorists	Non-decision theorists
Accept: one box	7	102
Lean toward: one box	1	88
Accept: two boxes	13	178
Lean toward: two boxes	6	95
Other	4	437
Total	31	900

Setting aside those who did not accept nor lean towards either one-boxing or two-boxing (i.e. those in the 'Other' row) 70.4 % of the decision theorists answered in favour of two-boxing whereas 59.0 % of the other philosophers answered in favour of two-boxing.

These numbers suggest that decision theorists are more prone to two-boxing than other philosophers.[5] My hypothesis is that the difference is due to ambiguities in Nozick's Newcomb problem and the existence of clearer Newcomb problems that are mostly known only to specialists. Nozick's initial Newcomb problem, which is probably the only one most philosophers who do not work in decision theory know, is unnecessarily obscure. Furthermore, since the nature of the being's predictive power is unclear it is unnecessarily obfuscated whether the contents of the boxes are causally independent of the agent's choice. As we will see below, there are other Newcomb problems, where the advantages of two-boxing (or the option corresponding to two-boxing) become more obvious.

The salient feature of a Newcomb problem is that there are two complementary states s_1 and s_2 and two alternatives a_1 and a_2 available to the agent such that:

- The agent knows that s_1 and s_2 are causally independent of a_1 and a_2.
- The agent's utilities are such that
  $U(a_1 \wedge s_1) > U(a_2 \wedge s_1)$ and $U(a_1 \wedge s_2) > U(a_2 \wedge s_2)$,
  where $U(x)$ is the agent's cardinal utility for x.
- The agent's utilities and subjective probabilities are such that
  $P(s_1|a_2)U(a_2 \wedge s_1) + P(s_2|a_2)U(a_2 \wedge s_2) > P(s_1|a_1)U(a_1 \wedge s_1) + P(s_2|a_1)U(a_1 \wedge s_2)$,
  where $P(x|y)$ is the agent's subjective probability for x conditioned on the evidence that y is chosen.

These features are more obvious in the following type of Newcomb problems, known as medical Newcomb problems:

> *The Smoking Lesion*
> Susan is debating whether or not to smoke. She believes that smoking is strongly cor-

[5] Note, however, that the difference is not statistically significant. Pearson's χ^2 test yields $p \approx 0.24$ and Fisher's exact test yields $p \approx 0.083$.

> related with lung cancer, but only because there is a common cause—a condition that tends to cause both smoking and cancer. Once we fix the presence or absence of this condition, there is no additional correlation between smoking and cancer. Susan prefers smoking without cancer to not smoking without cancer, and she prefers smoking with cancer to not smoking with cancer. Should Susan smoke? It seems clear that she should.[6]

A major advantage of the Smoking Lesion over Nozick's initial example is that it is less open to decision theoretically irrelevant misunderstandings. The only way you can achieve a better outcome in The Smoking Lesion, regardless of whether you have the gene, is to smoke. This makes the option corresponding to two-boxing, smoking, intuitively seem to be the only rational choice.

Newcomb problems are usually regarded as counterexamples to evidential decision theory as defended by, for example, Richard C. Jeffrey.[7] Evidential decision theory recommends deciding upon an option with maximum conditional expected utility:

> $\mathrm{VAL}_{\mathrm{EDT}}(x) = \sum_{s \in S} P(s|x)U(s \wedge x)$,
> where S is a partitioning of states of the world.
>
> *Evidential decision theory (EDT)*
> It is rational to decide upon an alternative x if, and only if, there is no other alternative with higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x.[8]

The trouble is that evidential decision theory recommends refraining from smoking in The Smoking Lesion. This recommendation is due to that EDT recommends options on account of their desirability as news. It would be good news to find out that you have chosen not to smoke since you are then likely not to have the gene. But it seems irrational to act as to get good news when you are not making the news.[9]

A common response to Newcomb problems is to reject EDT in favour of causal decision theory. David Lewis's version of causal decision theory recommends us to 'consider the expected value of your options under the several hypotheses; you should weight these by the

[6] Egan (2007, p. 94). The first occurrence of this problem in decision theory is due to Robert C. Stalnaker in a 1972 letter to David Lewis reprinted in Harper et al. (1981, p. 152). The problem was in all likelihood inspired by the views of Ronald A. Fisher (1957, p. 298) who suggested 'that cigarette-smoking and lung cancer, though not mutually causative, are both influenced by a common cause, in this case the individual genotype.'

[7] Jeffrey (1965).

[8] Here one might want to qualify 'rational' to 'rational given that the agent's desires and beliefs are rational'. Take for example Susan in The Smoking Lesion. Her beliefs might not be rational since she believes that smoking does not cause cancer in face of available evidence to the contrary. It might not be rational to choose based on irrational beliefs. The same qualification could be inserted into all decision theories discussed in this section.

[9] However, not everyone agrees that EDT does not recommend smoking. See e.g. Ellery Eells (1982).

credences you attach to the hypotheses; and you should maximise the weighted average.'[10] Let a dependence hypothesis be a maximally specific proposition about how the things the agent cares about depend causally on her options. Then, causal decision theory recommends choosing an option with maximum causal expected utility:

$$\mathrm{VAL}_{\mathrm{CDT}}(x) = \sum_{k \in K} P(k)U(k \wedge x),$$

where K is a partitioning of dependency hypotheses.

Causal decision theory (CDT)
It is rational to decide upon an alternative x if, and only if, there is no other alternative with higher $\mathrm{VAL}_{\mathrm{CDT}}$ than x.

For example, in The Smoking Lesion, Susan would compare smoking and refraining under the dependency thesis she is convinced of, namely that smoking causes enjoyment, not cancer. Thus, CDT only recommends smoking as rational in The Smoking Lesion.

2. Ratificationism

Jeffrey's initial response to the Newcomb problems was not to give up evidential decision theory completely, but to modify it with a requirement that one's decisions be ratifiable.[11]

> Ratificationism requires performance of the chosen act, *A*, to have at least as high an estimated desirability as any of the alternative performances *on the hypothesis that one's final decision will be to perform A.*[12]

The idea is that an option should have at least as high unconditional expected utility as any other option on the supposition that it is decided upon, where unconditional expected utility is defined as follows:

$$\mathrm{VAL}(x) = \sum_{s \in S} P(s)U(s \wedge x),$$

where S is a partitioning of states of the world.

An option x is *ratifiable* if, and only if, there is no alternative y such that $\mathrm{VAL}(y)$ exceeds $\mathrm{VAL}(x)$ on the supposition that x is decided upon.[13]

Then we can state Jeffrey's version of ratificationism:

[10] Lewis (1981, pp. 11–12).
[11] Jeffrey (1983, p. 16).
[12] Jeffrey (1983, p. 19).
[13] Egan (2007, p. 107).

> *Jeffrey's ratificationism (JR)*
> It is rational to decide upon an option x if, and only if, x is the only ratifiable option.

For an example, consider again The Smoking Lesion. Suppose that you decide not to smoke. This is good news—it is likely that you do have the gene that causes cancer. However, to smoke would have a higher VAL than not smoking since you prefer smoking whether or not you have the gene. Your decision not to smoke is hence not ratifiable. Since smoking is the only ratifiable option in The Smoking Lesion, JR recommends smoking.

A peculiar feature of Jeffrey's proposal is the requirement that there should only be one ratifiable option. He is fully aware that there are situations where no option is ratifiable and situations where more than one option is ratifiable. He mentions two such cases:

> *The green-eyed monster*
> Where the agent must choose one of two goods and see the other go to someone else, greed and envy may conspire to make him rue either choice. Decision theory cannot cure this condition, and ratificationism recommends neither option.[14]

> *The triumph of the will*
> A madly complacent agent could find all the acts ratifiable because with him, choice of an act always greatly magnifies his estimate of its desirability—not by changing probabilities of conditions, but by adding a large increment to each entry in the chosen act's row of the desirability matrix.[15]

In such situations, where there are either none or two or more ratifiable options, Jeffrey recommends you to 'reassess your beliefs and desires before choosing.'[16] However, while there is arguably something irrational about the agent's desires in the two above examples, there are other cases with none or two or more ratifiable options where the agent's beliefs and desires do not seem irrational. For such a case with two ratifiable options, consider a variation of The Smoking Lesion where there are two smoking options: smoke Dunhill or smoke Gauloises. Furthermore, suppose that Susan enjoys smoking both brands but she is indifferent between them. In this case both smoking options are ratifiable and Susan's beliefs and desires do not seem to be irrational.

For examples where no option is ratifiable while the agent's desires and beliefs do not seem irrational, consider Allan Gibbard and William L. Harper's Death in Damascus case or Andy Egan's The Psychopath Button:[17]

[14] Jeffrey (1983, p. 18).
[15] Jeffrey (1983, p. 19).
[16] Jeffrey (1983, p. 19).
[17] Gibbard and Harper (1978, pp. 157–158). The Death in Damascus case is quoted in Essay IV.

The Psychopath Button
Paul is debating whether to press the "kill all psychopaths" button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button. Paul very strongly prefers living in a world *with* psychopaths to dying.[18]

In this cases it seems irrational to press and rational to refrain from pressing.[19] It seems irrational to choose an option that you think, under the supposition that you decide upon it, is likely to cause the worst outcome. Since no option is ratifiable, JR recommends neither option.

The Psychopath Button also spells trouble for CDT which only recommends pressing the button. This is due to CDT only using the agent's unconditional credences for dependency theses. The unconditional credence for the hypothesis that pressing causes your death should be low given that your credence for not being a psychopath is high. CDT does not take into account that you have a high credence for that pressing the button causes your death conditional on you having decided to press—which seems to matter.

As we have seen, a problem with ratificationism is that in some cases there are no ratifiable options, but some options still seem rational in these cases. To remedy this problem, Egan tentatively considers a lexical version of ratificationism:

Lexical ratificationism (LR)
It is rational to decide upon an option x if, and only if,

1. x is ratifiable and there is no other ratifiable option with higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x, *or*
2. there are no ratifiable options, and no other (unratifiable) option has higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x.[20]

LR will recommend at least one option even in cases where no option is ratifiable. For example, it yields the intuitively right recommendation, not pressing, in The Psychopath Button. Nevertheless, LR goes wrong in the following case due to Anil Gupta:

The Three-Option Smoking Lesion
Samantha has three options: Smoke cigars, smoke cigarettes, or refrain from smoking

[18] Egan (2007, p. 97). The problem was suggested by David Brandon-Mitchell. It was probably inspired by Egan's similarly structured—but less catchy—The Murder Lesion, Egan (2007, p. 97).

[19] This reaction is, however, not entirely universal. John Cantwell (2010), who favours pressing, objects to Egan's diagnosis of the alleged irrationality of causal decision theory. James M. Joyce (forthcoming) also rejects the intuition that refraining is the only rational choice. Furthermore, Joyce disputes that CDT recommends pressing the button as the only rational choice.

[20] Egan (2007, p. 111).

> altogether. Call these options CIGAR, CIGARETTE, and NO SMOKE. Due to the ways that various lesions tend to be distributed, it turns out that cigar smokers tend to be worse off than they would be if they were smoking cigarettes, but better off than they would be if they refrained from smoking altogether. Similarly, cigarette smokers tend to be worse off than they would be smoking cigars, but better off than they would be refraining from smoking altogether. Finally, nonsmokers tend to be best off refraining from smoking.[21]

The only ratifiable option is NO SMOKE and, hence, LR's recommendation. But it seems strange to rule out CIGAR or CIGARETTE in favour of NO SMOKE due to their unratifiability since if you decide upon CIGAR or CIGARETTE then it is very likely that NO SMOKE would be your worst option.

Egan takes Gupta's example to be a counterexample, not just to LR but to every form of ratificationism.[22] I would not go that far. In fact I present a weakened version of ratificationism in Essay I that does not go wrong in Gupta's case, and that moreover yields the intuitively right recommendations in the other problem cases we have considered.[23]

2.1 *Some previous weakenings of ratifiability*

Before I present my proposal we will take a look at some previous weakenings of ratifiability by Paul Weirich and Wlodek Rabinowicz and assess whether they are adequate. Weirich introduces the concept of weak ratifiability.[24] In order to define weak ratifiability we first need some new terminology.

> A *path* from option x to option y is a sequence of options starting with x and ending with y such that for each option z in the sequence except for y the VAL of z on the supposition that z is decided upon is not higher than the VAL of the next option in the sequence on the supposition that z is decided upon.

Roughly, there is a path from x to y if you may reach a decision on y after tentatively deciding upon x and having then repeatedly revised your choice in light of your latest tentative decision.

> An option x is *opposed* to an option y if, and only if, the VAL of x on the supposition that y is decided upon is higher than the VAL of y on the supposition that y is decided upon, and there is no path from x back to y.[25]

[21] Egan (2007, p. 112).
[22] Egan (2007, p. 112).
[23] Gustafsson (forthcoming-b).
[24] Weirich (1986) and Weirich (1988).
[25] Weirich (1986, pp. 444–445).

> An option x is *weakly ratifiable* if, and only if, no option is opposed to x.[26]

Given weak ratifiability we can state the following weakening of ratificationism:

> *Weak ratificationism*
> If the states are known to be causally independent of the options, it is rational to decide upon an option x if, and only if, x is weakly ratifiable and there is no other weakly ratifiable option with higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x.[27]

However, weak ratificationism is not fully adequate. Rabinowicz has shown that it violates the causal version of the dominance principle:

> *Dominance with causal independence*
> If the states are known to be causally independent of the options it is not rational to decide upon an option x if there is an option y such that there is at least one positively probable state where the outcome of y is strictly preferred to the outcome of x and no state where the outcome of y is not weakly preferred to the outcome of x.

Rabinowicz found the following type of case, where the states are probabilistically dependent, but causally independent, of the options. For each $i \in \{1, 2, 3, 4\}$, the agent would consider her deciding upon a_i as a reliable sign that the world is in state s_i: [28]

	s_1	s_2	s_3	s_4
a_1	1	9	1	9
a_2	0	8	0	8
a_3	4	4	4	4
a_4	6	0	0	0

In this case only a_2 and a_3 are weakly ratifiable since a_1 and a_4 are opposed by a_3. Given that the agent would consider her deciding upon a_i as a very reliable sign that the world is in state s_i, then $\mathrm{VAL}_{\mathrm{EDT}}(a_2) > \mathrm{VAL}_{\mathrm{EDT}}(a_3)$. Hence, weak ratificationism recommends a_2 as the only rational choice. The trouble is that the utility of a_1 is higher than that of a_2 for every state. Thus, weak ratificationsim violates the dominance principle with causal independence.

In order to state Rabinowicz's weakened version of ratificationism we once again need some new terminology.

[26] Weirich (1988, p. 579).

[27] I here follow the presentation in Rabinowicz (1989, p. 628). As Rabinowicz notes, Weirich's proposal is slightly more complicated. However, the simplification will not matter for the objections we will consider.

[28] Rabinowicz (1989, p. 630).

> An option x is a *trap* with respect to an option y if, and only if, there is a path from y to x but not from x to y.
>
> An option x is *retrievable* if, and only if, no option is a trap with respect to x.[29]

We can then state Rabinowicz's proposal as follows: [30]

> *Retrievable maximization of expected utility (RMEU)*
> If states are known to be causally independent of the options, it is rational to decide upon an option x if, and only if, x is retrievable and there is no option with higher VAL than x.

RMEU does not violate the dominance principle with causal independence. Furthermore, it recommends smoking in The Smoking Lesion and if it does not recommend CIGAR or CIGARETTE in The Three-Option Smoking Lesion then that recommendation is due to a low unconditional expected utility and not to ratifiability or retrievability since all three options are retrievable. Also, there will always be at least one retrievable option. So far so good. However, in The Psychopath Button both pressing and not pressing the button are retrievable and due to the higher unconditional expected utility, RMEU recommends pressing the button. As mentioned earlier this recommendation seems counter-intuitive.

At this point one might object that this problem is easily fixed: The problem with RMEU that yields the wrong recommendation in The Psychopath Button is just that it chooses among the retrievable options by unconditional expected utility. What if we replace the unconditional expected utility with conditional?

> *Retrievable maximization of conditional expected utility (RMCEU)*
> If states are known to be causally independent of the options, it is rational to decide upon an option x if, and only if, x is retrievable and there is no option with higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x.

RMCEU recommends not pressing the button in The Psychopath Button. However, this improvement came at a very high price: RMCEU violates the dominance principle with causal independence. To see this, consider the following type of case where again, the states are known to be causally independent of the options and for each $i \in \{1, 2, 3\}$, the agent would consider her deciding upon a_i as a reliable sign that the world is in state s_i:

[29] Rabinowicz (1989, p. 637).

[30] Rabinowicz (1989, p. 638). Note that VAL is here restricted to states that are causally independent of the options, that is, it is here calculated as $\sum_{s \in S} P(s|x)U(s \wedge x)$, where S is a partitioning of states of the world that are causally independent of the options.

	s_1	s_2	s_3
a_1	2	9	3
a_2	1	7	2
a_3	3	8	1

All three options are retrievable. Furthermore, given that the agent takes her decision upon an option a_i as a reliable enough sign for the world to be in state s_i, a_2 will have the highest conditional expected utility. Hence, RMCEU only recommends choosing a_2. But since the utility of a_1 is higher than that of a_2 for every state, RMCEU violates the dominance principle with causal independence.

2.2 *General ratifiability*

Since none of the previous weakened versions of ratificationism yields the intuitively right recommendations in all the discussed cases, there is room for improvement. Essay I presents a new proposal, based on the concept of general ratifiability:

> An option x is *generally ratifiable* if, and only if, there is no option y such that for every option z, $\mathrm{VAL}(y)$ exceeds $\mathrm{VAL}(x)$ on the supposition that z is decided upon.

The intuition behind demanding that options should be generally ratifiable, is that if you predict that x will look better than y given that you choose any one of your available options then y does not seem like the way to go if x is available. My first tentative proposal in Essay I is then:

> *General ratificationism (GR)*
> It is rational to decide upon an option x if, and only if, x is generally ratifiable and there is no other generally ratifiable option with higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x.

Nevertheless, GR is not fully satisfactory. My worries are due to the following type of cases presented to me by Frank Arntzenius, where as before, the agent would consider her deciding upon a_i as a reliable sign that the world is in state s_i for each $i \in \{1, 2, 3\}$:

> *Scenario 1*

	s_1	s_2	s_3
a_1	2	4	1
a_2	1	3	2

Scenario 2

	s_1	s_2	s_3
a_1	2	4	1
a_2	1	3	2
a_3	0	0	0

Scenario 2 is like Scenario 1 except for the addition of the clearly dominated option a_3. The trouble is that GR recommends a_1 in Scenario 1 but a_2 in Scenario 2. The addition of the dominated a_3 should not make a difference for the choice between a_1 and a_2. In Scenario 1 the only generally ratifiable option is a_1, and thus GR's recommendation. In Scenario 2 both a_1 and a_2 are generally ratifiable, since a_2 has a higher $\mathrm{VAL}_{\mathrm{EDT}}$ than a_1 on the supposition that the agent decides upon the not generally ratifiable a_3. My diagnosis is that GR correctly rules out a_2 in Scenario 1 and a_3 in Scenario 2, but that the test for general ratifiability should have been repeated in Scenario 2 to rule out a_2 as not generally ratifiable in the choice between the remaining options a_1 and a_2.

In response to this problem Essay I offers another proposal, where the test for general ratifiability is repeated on the options that survived the previous test.

> An option x is *generally ratifiable*$_0$ if, and only if, there is no option y such that for every option z, $\mathrm{VAL}(y)$ exceeds $\mathrm{VAL}(x)$ on the supposition that z is decided upon.
>
> An option x is *generally ratifiable*$_{n+1}$ if, and only if, there is no generally ratifiable$_n$ option y such that for every generally ratifiable$_n$ option z, $\mathrm{VAL}(y)$ exceeds $\mathrm{VAL}(x)$ on the supposition that z is decided upon.
>
> An option x is *iteratively generally ratifiable* if, and only if, for all $k \geq 0$, x is generally ratifiable$_k$.
>
> *Iterated general ratificationism (IGR)*
> It is rational to decide upon an option x if, and only if, x is iteratively generally ratifiable and there is no other iteratively generally ratifiable option with higher $\mathrm{VAL}_{\mathrm{EDT}}$ than x.

Since the test for general ratifiability is repeated, a_2 is ruled out in the second iteration in Scenario 2. Thus, IGR recommends a_1 in both scenarios.

Furthermore, IGR yields the intuitively right recommendations in all the problem cases above. IGR recommends two-boxing in Newcomb problems. For example, in The Smoking Lesion only the decision to smoke is generally ratifiable since smoking has a higher VAL both on the supposition that smoking is decided upon and that non-smoking is chosen. Thus, IGR recommends smoking in The Smoking Lesion.

It is proven in Essay I that given a finite set of available options there will always be at least one generally ratifiable option. Hence, IGR will make a recommendation also in cases where no option is ratifiable e.g. The Green-Eyed Monster, Death in Damascus, and The Psychopath Button. In The Psychopath Button both pressing and not pressing the button are iteratively generally ratifiable. IGR recommends not pressing due to a higher $\mathrm{VAL}_{\mathrm{EDT}}$ for not pressing than for pressing. This is because under the supposition that you decide to press it is likely that you are a psychopath and hence that you die if you press, which is worse than the likely scenario if you do not press under the supposition that you decide not to press, that is, living in a world with psychopaths.

While the two smoking options in The Three-Option Smoking Lesion might be ruled out in favour of not smoking, it would not be due to them not being ratifiable (or for that matter generally ratifiable). All three options are iteratively generally ratifiable. Thus, if CIGAR or CIGARETTE are ruled out in favour of NO SMOKE, this will be due to a higher $\mathrm{VAL}_{\mathrm{EDT}}$ of NO SMOKE.

Thus, unlike previous decision theories, IGR gives the intuitively right recommendations in all the discussed problem cases.

3. Money pumps

In our discussion of decision theories so far, we have taken some rationality constraints for granted. Two of the most discussed of these rationality constraints are transitivity and completeness. This section will discuss the first of these constraints, transitivity. The other constraint, completeness, will be discussed in sections 4 and 5. All the decision theories discussed so far demand that transitive preferences are a prerequisite of rationality. Let 'xPy' denote that x is preferred to y and let 'xIy' denote indifference between x and y. Then, two transitivity principles, both required by classical decision theory, can be stated as follows: [31]

PP-transitivity
$\forall x \forall y \forall z((xPy \wedge yPz) \rightarrow xPz)$.

PI-transitivity
$\forall x \forall y \forall z((xPy \wedge yIz) \rightarrow xPz)$.

[31] Given completeness of weak preference *PP*- and *PI*-transitivity imply transitivity of weak preference. See Sen (1970, pp. 18–19) for proof. Therefore under completeness there is no need to defend other transitivity requirements like, for example, the transitivity of indifference since these will follow from *PP*- and *PI*-transitivity. Without completeness, on the other hand, one would have to defend requirements like transitivity of indifference independently. However, as I will argue, without completeness the situation is much worse since money-pump arguments do not work without completeness.

The standard argument for the claim that transitivity is rationally required is the money-pump argument. The money-pump argument purports to show that an agent who has intransitive preferences will in some possible situations be forced to act against her preferences. The first occurrence of a version of the money-pump argument in print is due to Donald Davidson, J. C. C. McKinsey, and Patrick Suppes (1955) who attribute it to Norman Dalkey.[32] It is part of a defence of *PP*-transitivity against a counterexample where a Mr. S considers three different jobs:

> a = full professor with a salary of $5,000.
>
> b = associate professor at $5,500.
>
> c = assistant professor at $6,000.

Mr. S holds the preferences aPb, bPc, and cPa. Davidson et al. object that these preferences rule out a rational choice according to the following principle:

> a rational choice (relative to a given set of alternatives and preferences) is one which selects the alternative which is preferred to all other alternatives; if there are several equivalent alternatives to which none is preferred, then any of these is selected.[33]

This principle is then summarized to:

> *Non-dominated choice*
> a rational choice is one which selects an alternative to which none is preferred.[34]

However, non-dominated choice differs from the longer principle in that it may grant as rational an alternative that is incommensurate with another alternative. If neither preference in either direction nor indifference may hold between alternatives, this seems like a welcome feature. To illustrate the non-dominated choice principle Davidson et al. introduce the money pump.

> We may imagine a scene in which the point becomes obvious. The department head, advised of Mr. S's preferences, says, 'I see you prefer b to c, so I will let you have the associate professorship—for a small consideration. The difference must be worth something to you.' Mr. S. agrees to slip the department head $25. to get the preferred alternative. Now the department head says, 'Since you prefer a to b, I'm prepared—if you will pay me a little for my trouble—to let you have the full professorship.' Mr. S. hands over another $25. and starts to walk away, well satisfied, we may suppose. 'Hold on,' says the department head, 'I just realized you'd rather have c than a. And I can arrange that—provided…' [35]

[32] Davidson et al. (1955, p. 146, fn. 4).
[33] Davidson et al. (1955, p. 145).
[34] Davidson et al. (1955, p. 145).
[35] Davidson et al. (1955, p. 146).

Since the example is supposed to illustrate the non-dominated choice principle it seems that the purported irrationality of Mr. S is supposed to be that he is never satisfied with his decision; he always wants to pay to swap to another alternative. Another reason to judge Mr. S to be irrational is that he is forced to act against his preferences through a series of steps by accepting the department head's offers.

A weakness in their defence of *PP*-transitivity from the counterexample with Mr. S is that one can vary the example by introducing a Mrs. T who holds the preferences aPb, bPc, and cIa.[36] Mrs. T's preferences also violate *PP*-transitivity (and furthermore, *PI*-transitivity). Nevertheless, she can make a choice that is not ruled out as irrational by the non-dominated choice principle. This is because she does not prefer any alternative to *a*. Similarly, Mrs. T cannot be exploited in the same way as Mr. S since she does not rather have *c* than *a*, she has no reason to accept the department head's offer of *c* for *a*.

3.1 *Forcing and non-forcing money pumps*

So Mrs. T does not violate the non-dominated choice principle and she does not need to accept a swap from *c* to *a*. But since she is indifferent between *a* and *c* she might, without acting contrary to her preferences, swap from *c* to *a*. One might hold that it cannot be rationally permitted to be money pumped. Thus, one might charge Mrs. T with irrationality since her preferences allows her to go along with the department head's scheme.

Essay III makes a distinction between forcing and non-forcing money pumps.[37] A forcing money pump is a money pump like the one employed against Mr. S where the agent must either accept every swap or choose against his preferences. A non-forcing money pump is a money pump where in at least one step the agent may either accept the swap or she might reject it, without choosing contrary to her preferences.[38]

While Mrs. T is not susceptible to the standard forcing pump employed by Davidson et al. above, she is susceptible to a non-forcing pump since she might accept a swap from *c* to *a* if the department head offered it without any fee. If she accepted a free swap from *c* to *a* she may still be pumped for money if she paid the department head for the other swaps. The idea here is that it should not be rationally permitted to go along with a money pump, and Mrs. T may do so without acting contrary to her preferences.

[36] Nozick (1963, p. 88).

[37] Gustafsson and Espinoza (2010, pp. 761–762).

[38] Cf. Sven Ove Hansson's (1993, pp. 478–479) distinction between P^*-pumps, that pump cycles of strict preferences, and R^*P-pumps, that pump cycles of weak preference with at least one strict preference. Hansson's distinction concerns the types of preferences that are pumped. It turns on whether the preferences that are pumped are cycles of strict preference. The distinction between forcing and non-forcing pumps, on the other hand, concerns whether the agent is required to go along with the pump. The distinctions come apart; the upshot of Essay II is that there are both forcing P^*-pumps and forcing R^*P-pumps.

However, I argue in Essay II that money-pump arguments are not cogent if they rely on non-forcing money pumps.[39] The crucial difference between forcing and non-forcing money pumps is that in a forcing money pump the agent either lets himself be pumped or acts contrary to his preference in some step, while in a non-forcing money pump the agent may avoid being pumped for money without acting contrary to her preferences. The problem is that even though Mrs. T is indifferent between *a* and *c*, she may still be rationally forbidden to swap *c* for *a* because of some other rationality constraint. In order for the non-forcing version of the money-pump argument to get off the ground, one would have to show first, without begging the question, that there are no rationality constraints that forbid Mrs. T from swapping *c* for *a*. But the prospects for this endeavour look dim.

3.2 *A forcing money pump for intransitive preferences*

The standard approach for making agents like Mrs. T susceptible to a forcing money pump is to offer the agent a small premium for the swaps between alternatives she is indifferent.[40] The idea is that if you are indifferent between *a* and *c* then you should prefer *a* with a small monetary premium to *c*. As long as the agent pays more for the swap where she has a preference for the other alternative she will still be money pumped.

However, this standard approach is shown to be question begging in Essay II. In short, the charge is that in order to conclude that the agent prefers *a* with a premium to *c* just because she is indifferent between *a* and *c* one needs to invoke transitivity of preference. And to rely on that transitivity is rationally required in an argument that transitivity is rationally required is to beg the question.

Essay II presents a new approach that does not rely on transaction premiums.[41] This makes use of the plausible dominance principle: [42]

> *Dominance*
> If there is a partition of states of the world such that it is independent of lotteries L' and L'' and relative to it, there is at least one positively probable state where the outcome of L' is strictly preferred to the outcome of L'' and no state where the outcome of L' is not weakly preferred to the outcome of L'', then L' is strictly preferred to L''.

Except for the debate due to Newcomb's problem (see Section 1) on whether the independence of the lotteries should be causal or evidential, the dominance principle is relatively

[39] Gustafsson (2010b, p. 252).
[40] See, e.g. McClennen (1990, pp. 90–91).
[41] Gustafsson (2010b, pp. 255–256).
[42] Savage (1951, p. 58).

uncontroversial. Fortunately, the cogency of my argument does not hinge on what type of independence is required.

The approach works as follows: If an agent satisfies both completeness (that is, for any pair of alternatives she is either indifferent between them or she prefers one to the other) and dominance, but her preferences over the alternatives a, b, and c violate transitivity (PP or PI) then she has cyclical preferences over the following lotteries:

	S_1	S_2	S_3
L_1	a	b	c
L_2	b	c	a
L_3	c	a	b

Here the states S_1, S_2, and S_3 have been chosen such that they are independent (in the way required by the dominance principle) from the lotteries L_1, L_2, and L_3. For example, both Mr. T and Mrs. S will either violate dominance or have the cyclic preferences L_1PL_2, L_2PL_3, and L_3PL_1. Given cyclic preferences over these lotteries one can then employ the standard money pump sketched by Davidson et al.

3.3 *The irrelevance of exploitability and resolute choosers*

Frederic Schick has levelled an influential objection to money-pump arguments. He argues as follows:

> Again, the agent prefers C to B, B to A, and A to C. This much remains fixed. It does not follow that the values he sets on the arrangements he is offered are all positive. In the absence of special information, he sets a positive value on the pumper's canceling X in favor of some preferred outcome Y—this for all X and Y. But where he has made certain arrangements already and now looks back, he may get the drift. He may see he is being pumped and refuse to pay for any further deals. His values would then be different. He would set a *zero* value on any new arrangement.[43]

Schick's point seems to be that the agent may prefer a to c but he may prefer

> c after having swapped a for b and then b for c

to

> a after having swapped a for b, b for c, and then c for a.

[43] Schick (1986, p. 118).

Even though an agent has cyclical preferences over some alternatives the agent's preferences may be different for the combination of the alternatives and a sequence of swaps. The alternatives may not be preference-wise independent.[44] Thus, one may have cyclic preferences and still turn down the second or third swap in the money pump.

A similar objection is due to Edward F. McClennen. He proposes that one may avoid being money pumped by becoming a resolute chooser. A resolute chooser is someone who

> […] proceeds, against the background of his decision to adopt a particular plan, to do what the plan calls upon him to do, even though it is true (and he knows it to be true) that were he not committed to choosing in accordance with that plan, he would now be disposed to do something quite distinct from what the plan calls upon him to do.[45]

If one does not confront each new decision myopically, but instead adopt and stick to a plan one may avoid being money pumped. For example, Mr. S may adopt the plan to accept the swap from *a* to *b* and also from *b* to *c* and then refuse any further trades.[46] Hence, Mr. S could avoid being money pumped.

Before we reply to these objections, we need to differentiate between two views on what is supposed to be irrational about the agent who goes along with a money pump. Schick takes a premise of the money-pump argument to be that it is irrational to be exploited. He writes about jointly exploitable dispositions, 'My point has been only that their being exploitable does not reveal any fault in them.' But on my view it is not being exploitable by itself that is irrational. What is irrational about being money pumped is that one chooses against one's preferences. For example, choosing *a* with a loss of money over *a* without a loss of money, when you prefer *a* without a loss of money to *a* with a loss of money. Whether someone else thereby gets rich on your expense is irrelevant for whether you are rational. If you do not mind being exploited then the classical decision theorist may grant your letting yourself be exploited as rational.

Note that there is no talk of exploitability in the original presentation of the money-pump argument by Davidson et al. Their point does not seem to be that Mr. S is irrational because he is exploited by the department head. The money-pump example is supposed to illustrate the non-dominated choice principle which yields that it is irrational to choose an alternative to which another alternative is preferred. It is that Mr. S is not satisfied whatever he chooses that is supposed to be irrational—he is always willing to pay in order to revoke his decision in favour of another alternative.

---

[44] Schick (1986, p. 118) writes 'value-wise independent' but this is confusing since we are dealing with preferences, not values nor value judgements.

[45] McClennen (1990, p. 13).

[46] McClennen (1990, p. 166).

Since it is not exploitability but acting against one's preference that is taken to be irrational, the sequential part of the argument is unnecessary. The department head could offer Mr. S a choice between all three of a, b, and c, at once. This will not make Mr. S poor nor the department head rich, but it will force Mr. S to choose an alternative over which another is preferred which the non-dominated choice principle rules out as irrational.

Since Mr. S this time just makes a single choice between the alternatives individually it does not matter if the alternatives are not preference-wise independent. Thus, Schick's worry is irrelevant. Furthermore, in reply to McClennen, since Mr. S in this variation only makes one choice, any plans are irrelevant. Once again, in order to use the same tactic against Mrs. T one can employ the approach presented in Essay II to elicit strict cyclic preferences over some lotteries and then offer her a choice between all of them.

Here one might object that it is more worrying to be ruined by exploitation than just acting against one's preference. Thus, some of the punch of the sequential money-pump argument is lost in the non-sequential one. But, even though the prospect of losing all one's money makes the argument more dramatic, what is supposed to be irrational about losing all one's money? It just seems irrational since most people prefer not to be ruined, and thus to choose to be ruined when given the choice is to choose against one's interests. Thus, the non-sequential version of the argument should be equally worrying, since it involves the same type of irrationality.

3.4 *A money pump for incomplete preferences?*

The careful reader noted that I assumed that the agents satisfied completeness in my version of the money-pump argument. This is a source of worry since, as we will see in sections 4–7, completeness has been challenged. For two examples we introduce a new couple: Mr. X and Mrs. Y. Let '$\diagup$' denote preferential incomparability. Mr. X holds the preferences aPb, bPc, and $a \diagup c$. Mrs. Y holds the preferences aPb, bIc, and $a \diagup c$. Mr. X violates PP-transitivity and Mrs. Y violates PI-transitivity. If completeness is not rationally required then one also needs to show why these violations of transitivity are irrational. The problem is how Mr. X and Mrs. Y could be made susceptible to a money pump if completeness is not rationally required.

None of the methods employed above against Mr. S and Mrs. T work on Mr. X and Mrs. Y The standard money pump that was sketched by Davidson et al. does not work since Mr. X and Mrs. Y do not need to have cyclic preferences. The dominance based approach presented in Essay II does not work here since Mr. X and Mrs. Y do not weakly prefer a to c nor vice versa. Nevertheless, some money pumps that can pump incomplete preferences have been

proposed.[47] However, these money pumps have all been of the non-forcing type and as I argue in Essay II, non-forcing money pumps are unconvincing.

4. The small-improvement argument

Like transitivity, completeness is also one of the most discussed assumptions of classical decision theory. As mentioned in Section 3, completeness is the claim that for any pair of alternatives an agent is rationally required to either prefer one of the alternatives to the other or to be indifferent between them. Thus, completeness requires that one of the traditional three preference relations holds between any pair of alternatives. More formally the condition can be stated as follows:

> *Completeness*
> $\forall x \forall y (xPy \vee yPx \vee xIy)$.

This section will present the most influential argument against completeness, namely the small-improvement argument in its various versions. Essay III argues against this argument. As will be described below there are some modified versions of the argument that escape the criticism put forward in Essay III, that Erik Carlson has brought to my attention. For my present views on why the small-improvement argument is unsuccessful (also in these modified versions), see Section 5.

The small-improvement argument was first proposed by Ronald de Sousa under the title 'the case of the Fairly Virtuous Wife'. He writes:

> I tempt her to come away with me and spend an adulterous weekend in Cayucos, California. Imagine for simplicity of argument that my charm leaves her cold. The only inducement that makes her hesitate is money. I offer $1,000 and she hesitates. Indeed she is so thoroughly hesitant that the classical decision theorist must conclude that she is *indifferent* between keeping her virtue for nothing and losing it in Cayucos for $1,000. […] The obvious thing for me to do now is to get her to the point of clear preference. That should be easy: everyone prefers $1,500 to $1,000, and since she is indifferent between virtue and $1,000, she must prefer $1,500 to virtue by exactly the same margin as she prefers $1,500 to $1,000: or so the axioms of preference dictate. Yet she does not. As it turns out she is again 'indifferent' between the two alternatives. The classical Utilitarian is forced to say that she is incoherent, because she violates his axioms of rationality. […] I would prefer to say that the alternatives considered are *incomparable*. […] We have dropped connexity, but she is not irrational.[48]

[47] Peterson (2007).
[48] de Sousa (1974, pp. 544–545).

In de Sousa's original rendition the argument is purely about rational preferences. The common structure of all versions of the small-improvement argument is as follows: first we have a premise about some kind of comparisons, and then we have some kind of transitivity premise from which it follows that none of the traditional comparative relations holds. The preferential version can be stated formally as follows:

The small-improvement argument (original preferential version)

(P1) A set of rational preferences satisfies
$\exists x \exists y \exists z (\neg(xPy) \wedge \neg(yPx) \wedge zPx \wedge \neg(zPy))$.

(P2) All rational preferences satisfy
$\forall x \forall y \forall z ((xPy \wedge yIz) \rightarrow xPz)$.

(P3) A set of rational preferences satisfies
$\exists x \exists y \neg(xPy \vee yPx \vee xIy)$.

Essay III argues that this original version of the argument suffers from a conflict between the reasons to believe the two premises. The argument does not satisfy the following condition:

> *Assumption of other conjuncts*
> A collection of reasons to believe the individual conjuncts of a conjunction provides a reason to believe the conjunction only if they are reasons to believe each conjunct under the assumption that the other conjuncts are true.[49]

The problem is that if one assumes (P1), then the money-pump argument cannot support (P2), since we then would have to allow for non-completeness. As explained in Section 3 the money-pump argument does not work if one cannot rule out non-completeness. In a reply to Essay III, Carlson shows that a revised version of the preferential small-improvement argument does not have a conflict between the reasons to believe its premises.[50] Carlson replaces (P2) with a weaker premise:

[49] Gustafsson and Espinoza (2010, p. 758).
[50] Carlson (forthcoming).

The small-improvement argument (revised preferential version)

(P1) A set of rational preferences satisfies
$\exists x \exists y \exists z (\neg(xPy) \wedge \neg(yPx) \wedge zPx \wedge \neg(zPy))$.

(P2*) If all rational preferences satisfy
$\forall x \forall y (xPy \vee yPx \vee xIy)$,
then all rational preferences satisfy
$\forall x \forall y \forall z ((xPy \wedge yIz) \rightarrow xPz)$.

(P3) A set of rational preferences satisfies
$\exists x \exists y \neg(xPy \vee yPx \vee xIy)$.

The assumption of (P1) does not, in this revised version of the argument, conflict with the use of the money-pump argument to support (P2*). The important difference is that (P2*) only claims that transitivity is rationally required if completeness is required. Thus, one does not need to allow for the possibility of non-completeness when one supports (P2*) with the money-pump argument.

However, one does not need to invoke any transitivity principle at all. (P2*) can be replaced by the even weaker (P2**):

The small-improvement argument (minimal preferential version)

(P1) A set of rational preferences satisfies
$\exists x \exists y \exists z (\neg(xPy) \wedge \neg(yPx) \wedge zPx \wedge \neg(zPy))$.

(P2**) If a set of rational preferences satisfies
$\exists x \exists y \exists z (\neg(xPy) \wedge \neg(yPx) \wedge zPx \wedge \neg(zPy))$,
then a set of rational preferences satisfies
$\exists x \exists y \neg(xPy \vee yPx \vee xIy)$.

(P3) A set of rational preferences satisfies
$\exists x \exists y \neg(xPy \vee yPx \vee xIy)$.

The idea is that (P2**) can be supported by a money-pump argument directly, without any resort to transitivity. To see this, consider the virtuous wife in de Sousa's example. Let a be 'lose virtue for $1000', b be 'keep virtue', and c be 'lose virtue for $1500'. Then she has the preferences $\neg(aPb) \wedge \neg(bPa) \wedge cPa \wedge \neg(cPb)$. If she satisfies completeness she has one of the following preferences: $aIb \wedge cPa \wedge bIc$ or $aIb \wedge cPa \wedge bPc$. Then she can be money pumped with the method developed in Essay II. We introduce three lotteries L_1, L_2, and L_3 that pay as follows, where S_1, S_2, and S_3 are three states of nature such that they are independent of the lotteries and positively probable:

	S_1	S_2	S_3
L_1	a	b	c
L_2	b	c	a
L_3	c	a	b

Then the virtuous wife will either violate the very plausible dominance principle or she has the exploitable cyclic preferences $L_1PL_2 \wedge L_2PL_3 \wedge L_3PL_1$.

Chang states the small-improvement argument in terms of value judgements.[51] Let '$\succ$' denote the relation *judged better than* and let '$\sim$' denote the relation *judged equally good as*. Then, the small-improvement argument for value judgements can be stated as follows:

The small-improvement argument (value judgement version)

(V1) A set of rational value judgements satisfies
$\exists x \exists y \exists z (\neg(x \succ y) \wedge \neg(y \succ x) \wedge z \succ x \wedge \neg(z \succ y))$.

(V2) All rational value judgements satisfy
$\forall x \forall y \forall z ((x \succ y \wedge y \sim z) \rightarrow x \succ z)$.

(V3) A set of rational value judgements satisfies
$\exists x \exists y \neg(x \succ y \vee y \succ x \vee x \sim y)$.

A stock objection to this version of the argument is that one cannot know for certain that the judgements involved in (V1) are right.[52] However, Chang argues that it is possible to find examples in which we have all the relevant knowledge to make the judgements in (V1) with certainty. In one such example one compares the tastes of coffee and tea:

> Suppose you must determine which of a cup of coffee and a cup of tea tastes better to you. The coffee has a full-bodied, sharp, pungent taste, and the tea has a warm, soothing fragrant taste. It is surely possible that you rationally judge that the cup of Sumatra Gold tastes neither better nor worse than the cup of Pearl Jasmine and that although a slightly more fragrant cup of the Jasmine would taste better than the original, the more fragrant Jasmine would not taste better than the cup of coffee.[53]

A possible problem with this move, to base the value judgements to subjective judgements of taste, is that (V2) becomes vulnerable to objections similar to those commonly raised against the transitivity of indifference. Suppose there are three cups of coffee: c_0 with no

[51] Chang (1997, pp. 23–24). See Sinnott-Armstrong (1985, p. 327) for a similar version in terms of moral requirements.

[52] See, e.g. Regan (1988, p. 1061).

[53] Chang (2002, p. 669).

sugar, c_1 with one lump of sugar, and c_2 with two lumps of sugar. An agent may judge c_0 and c_1 equally good because she cannot taste any difference between coffee with no sugar and coffee with merely one lump of sugar. Similarly, she may judge c_1 and c_2 equally good since she cannot taste the difference between coffee with one lump of sugar and coffee with two lumps. She might, however, be able to taste the difference between coffee with two lumps and coffee with no sugar at all, and therefore judge c_2 better than c_0.[54]

Let us now consider a purely axiological version of the argument. Surprisingly, this version has received little attention in the literature. Let 'B' denote the relation 'better than' and let 'E' denote the relation 'equally good as'. Then the axiological version of the small-improvement can be stated as:

The small-improvement argument (axiological version)

(A1) $\exists x \exists y \exists z (\neg(xBy) \wedge \neg(yBx) \wedge zBx \wedge \neg(zBy))$.

(A2) $\forall x \forall y \forall z ((xBy \wedge yEz) \rightarrow xBz)$.

(A3) $\exists x \exists y \neg(xBy \vee yBx \vee xEy)$.

Here the conclusion (A3) denies the axiological version of completeness: that one of the traditional value relations better, worse, or equally good holds between any pair of alternatives or more formally,

Axiological completeness
$\forall x \forall y (xBy \vee yBx \vee xEy)$.

The transitivity premise (A2) seems unproblematic. In this rendition of the argument the transitivity principle is more intuitively plausible than the corresponding transitivity principles (P2) and (V2). The trouble is that the same does not hold for premise (A1), which may also explain the lack of attention paid to this version in the literature.

Finally, there has been at least one attempt to construe the small-improvement argument in a radically different way. Sven Ove Hansson and Till Grüne-Yanoff write:

> When observing an agent choosing $C(\{X, Y\}) = \{X, Y\}$, the observer makes the agent repeat the choice, now with an offer of a small independent incentive i attached to one of the alternatives. If the agent chooses $C(\{X \wedge i, Y\}) = \{X \wedge i\}$, the observer may conclude that the agent was indifferent between X and Y, and that the addition of i to X shifted the balance to $X \wedge i$ over Y. If the agent however chooses $C(\{X \wedge i, Y\}) = \{X \wedge i, Y\}$, then the observer may conclude that X and Y were incomparable for the agent, and

[54] Cf. Luce (1956, p. 179).

> that the addition of i to X did not alter X's incomparability to Y. Because the agent's evaluation of i is presupposed, this method is not uncontroversial.[55]

Surely, this method is flawed but hardly for the reason Hansson and Grüne-Yanoff think. In my view the assumption criticized by Hansson and Grüne-Yanoff, that the agent prefers $x \wedge i$ to just x, could be avoided by observing whether the agent chooses $C(\{X \wedge i, X \wedge \neg i\}) = \{X \wedge i\}$. To make this observation would not be more problematic than making the observations Hansson and Grüne-Yanoff grant as unproblematic. However, a more critical flaw of the method is that there is no way of observing whether the agent finds X and Y incomparable or if she is merely violating transitivity.

5. Incomparability and indeterminacy

There has been a number of attempts to find flaws in the small-improvement argument. Most of the previous counter-arguments hinge on the idea that the small-improvement argument will not work if we make some plausible assumptions about indeterminacy. In this section I argue that these counter-arguments are unconvincing. Nevertheless, a suggestion due to Wlodek Rabinowicz is promising. However, Ruth Chang has levelled two objections against this type of proposal. I defend Rabinowicz's approach against Chang's objections. I defend the position that the comparative judgements in the cases appealed to in the small-improvement argument are indeterminate. Furthermore, if one cannot rule out that the judgements in the small-improvement argument are indeterminate then the argument does not work.

5.1 *The collapsing principle*

John Broome does not argue directly against the small-improvement argument. He, nevertheless, objects to putative counterexamples to completeness like those employed in the small-improvement argument. Broome argues that these alleged counterexamples are really just examples of indeterminacy. If he is right, the small-improvement argument must be flawed. However, his case depends on a controversial principle:

> *The collapsing principle, special version.* For any x and y, if it is false that y is *F*er than x and not false that x is *F*er than y, then it is true that x is *F*er than y.[56]

[55] Hansson and Grüne-Yanoff (2009).

[56] Broome (1997, p. 74). In Broome (2004, p. 174) he states the principle in terms of what one can deny:

> *Collapsing principle.* For any predicate *F* and any things *A* and *B*, if we can deny that *B* is *F*er than *A*, but we cannot deny that *A* is *F*er than *B*, then *A* is *F*er than *B*.

This principle has been subject to a number of counterexamples by Erik Carlson. The examples are all of essentially the same structure. The shortest one runs as follows:

> [S]uppose that *A* and *B* are two identical alarm clocks, except that *A* is waterproof, and *B* is not. Is *A* a better alarm clock than *B*? There may be no definite answer, since it may be indeterminate whether water resistance is a good-making characteristic of artefacts that are not very likely to come into contact with water. It is clear, however, that *B* is not better than *A*, since *A*'s being waterproof definitely does not detract from its goodness as an alarm clock.[57]

Broome, nevertheless, remains unconvinced. All of Carlson's examples trade on there being some kind of indeterminacy about value-making features. Broome rejects that whether a certain feature contributes to the value of an object could be indeterminate.[58] However, this answer does not work if we modify Carlson's examples so that it is determinate what features contribute to the goodness of an object but indeterminate whether the object has one of these features. Suppose that *A* and *B* are two prospective cavaliers, identical in every relevant aspect except, it is indeterminate whether *B* is bald but it is determinate that *A* is not bald. For superficial reasons baldness contributes negatively to one's goodness as a cavalier. Then, surely, *B* is not better than *A*. But since it is indeterminate whether *B* is bald, it is indeterminate whether *B* differs from *A* in any relevant respect that contributes negatively to *B*'s goodness. Thus, it should be indeterminate whether *A* is better than *B*.

Since it is determinate what features contribute to the goodness of an object in this example it is not blocked by Broome's answer to Carlson. Furthermore, it seems unappealing to claim that it cannot be indeterminate whether an object has a certain feature that contributes to its *F*ness.

I shall now argue that Broome's positive argument for the collapsing principle begs the question. Broome argues as follows:

> My only real argument is this: If it is false that *y* is *F*er than *x*, and not false that *x* is *F*er than *y* then *x* has a clear advantage over *y* in respect of its *F*ness. So it must be *F*er than *y*. It takes only the slightest asymmetry to make it the case that one thing is *F*er than another. One object is heavier than another if the scales tip ever so slightly toward it. Here there is a clear asymmetry between *x* and *y* in respect of their *F*ness. That is enough to determine that *x* is *F*er than *y*.[59]

The unpersuasive step is the inference from that it is false that *y* is *F*er than *x*, and not false that *x* is *F*er than *y* to that *x* has a clear advantage over *y* in respect of its *F*ness. Of course

[57] Carlson (2004, p. 224).

[58] Broome (2009, p. 417). Cf. Broome (2004, pp. 185–186) where he at least admits the examples as a strong objection.

[59] Broome (1997, p. 74).

we can infer that x has a clear F-related advantage over y, namely, it is either determinate or indeterminate whether x is Fer than y whereas it neither determinate nor indeterminate whether y is Fer than x. But that this clear F-related advantage should translate into a clear advantage of x over y with respect to Fness seems unfounded. It merely implies that either x has a clear advantage over y in respect of its Fness or it is merely indeterminate whether x has an advantage over y in respect of its Fness. And, of course, if it is only indeterminate whether x has an advantage of over y in respect of its Fness then it is not determinate that x is Fer than y. So, Broome's only real argument for the collapsing principle seems to beg the question.

Nevertheless, in order to reinforce the obviousness of his argument, Broome offers an accompanying example. In this *gedankenexperiment* you have to name a new Canberra suburb. The suburb should be named after the greatest Australian who does not yet have a suburb. You have narrowed down the candidates to the two Australians Exe and Wye. You have concluded after an investigation that it is false that Wye is a greater than Exe, but that it is not false that Exe is greater than Wye. Broome judges it to be quite wrong to give the suburb to Wye. The upshot is that unless Exe is the greatest Australian, it cannot be obvious that one should name the suburb after Exe.[60] Broome claims:

> When it is false that y is Fer than x but not false that x is Fer than y, then if you had to award a prize for Fness, it is plain you should give the prize to x. But it would not be plain unless x was Fer than y. Therefore, x is Fer than y. This must be so whether you actually have to give a prize or not, since whether or not you have to give a prize cannot affect whether or not x is Fer than y.[61]

Two replies: Firstly, one possibility is that it could be permissible to give the suburb to Exe but still indeterminate whether one should give him the suburb. If it is indeterminate whether one should award the price for Fness to x, then it would not be strange if it was indeterminate whether x is Fer than y.

Secondly, even if one grants that it is obvious that one should give the suburb to Exe, this obviousness may not be due to Exe being greater than Wye. If it is obvious that one should give the prize for Fness to x then this may be due to it being false that y is Fer than x and indeterminate whether x is Fer than y if one finds a rationality constraint like the following obvious:

> *Avoid indeterminate worseness (AIW)*
> If x is the only option determinately not worse than any other option, choose x.

[60] Broome (1997, pp. 74–75).
[61] Broome (1997, p. 75).

This principle seems to be supported by the same intuitions that Broome appeals to in his example. Nevertheless, with AIW one may accept the wrongness of giving the suburb to Wye without giving in to the collapsing principle.[62] Thus, Broome's attempted vindication of the collapsing principle does not succeed. Hence, his defence of completeness, which depends on the collapsing principle, is not cogent.

5.2 *The mere possibility of evaluative indeterminacy*

In a recent paper Nicolas Espinoza argues that the small-improvement argument fails due to the mere possibility of evaluative indeterminacy. He writes:

> Let the letter D stand for determinate truth and the letter I stand for indeterminate truth (where $I\alpha$ is equivalent to $\neg D\alpha \wedge \neg D\neg\alpha$). Also note the following logical property which is a trivial expansion of the law of excluded middle:
>
> (EM) $D\alpha$ if and only if $\neg(D\neg\alpha \vee I\alpha)$ (and of course, $\neg D\alpha$ if and only if $D\neg\alpha \vee I\alpha$)[63]

He then presents a version of the small-improvement argument that takes into account the distinction between 'determinate and indeterminate truthness'. It goes as follows, where B is the relation 'better than', E is the relation 'equally good as', x^+ is x with a small improvement, and '[Refs. n,m]' denotes that the preceding proposition is inferred from propositions n and m:

(1) $D\neg(xBy) \wedge D\neg(yBx)$ [J1*]
(2) $D(x^+Bx)$ [J2*]
(3) $[D(xEy) \wedge D(x^+Bx)] \rightarrow D(x^+By)$ [IP$^+$]
(4) $D\neg(x^+By)$ [J4$^+$]
(5) $D\neg[D(xEy) \wedge D(x^+Bx)]$ [Refs. 3,4]
(6) $\neg D(xEy)$ [Refs. 2,5]
(7) $D\neg(xBy) \wedge D\neg(yBx) \wedge \neg D(xEy)$ [Refs. 1,6].[64]

The trouble with (7) according to Espinoza is that it does not rule out that it is indeterminate that x and y are equally good. The third conjunct just states $\neg D(xEy)$ which according to (EM) is equivalent to $D\neg(xEy) \vee I(xEy)$. Espinoza argues that the small-improvement argument fails since it cannot rule out that

[62] The same reply can, *mutatis mutandis*, be given to Broome's similar example with Sartre's student in Broome (2004, pp. 172–174).

[63] Espinoza (2008, p. 131).

[64] Formulas (1), (2), (3), and (4) are premises. Espinoza (2008, p. 131), brackets in original. The argument would make more sense if (5) was replaced by

(5*) $\neg D(xEy) \vee \neg D(x^+Bx)$.

(xi) $D\neg(xBy) \wedge D\neg(yBx) \wedge I(xEy)$.[65]

However, Carlson objects that if axiological completeness holds, then the following equivalence is true:

D-trichotomy: $D(xEy) \Leftrightarrow D\neg(xBy) \wedge D\neg(yBx)$ [66]

Given that axiological completeness implies D-trichotomy, it is easily shown that if (xi) is true then axiological completeness is false. So Espinoza's argument is blocked.

To this Espinoza gives what I take to be an unsatisfactory reply. Espinoza declares that he will attempt to show that Carlson's D-trichotomy principle is false. His argument for this seems to be that 'There may be cases when it is neither true nor false that the comparison pair is coverable by the comparison predicate.' [67] But this is irrelevant since Carlson only claims that axiological completeness implies D-trichotomy, and therefore, only cases where it is true that all alternatives are comparable with respect to value (and thus coverable by the comparison predicates 'better' or 'equally good') are relevant as counterexamples. Since Espinoza has not been able to adequately answer Carlson's objection, his case against the small-improvement argument is unconvincing.

5.3 *Rabinowicz's analysis*

It does not seem to be a problem for the small-improvement argument that comparative judgements may be indeterminate as long as the judgements appealed to in argument are determinate. However, one might go further than Espinoza and question whether the judgements by, for example, de Sousa's virtuous wife and Chang's coffee and tea drinker are indeterminate. Indeed, this has been questioned by Rabinowicz. Rabinowicz has recently argued that the small-improvement argument loses its force if we grant that the judgements appealed to in the argument may be indeterminate. He writes:

> The introduction of x^+ does not allow us to definitely rule out the possibility of x and y being equally good, as long as we cannot definitely establish that x^+ is not better than y. The following are mutually compatible claims:
>
> (i) It is indeterminate whether x is equally as good as y.
> (ii) It is determinate that x^+ is better than x.
> (iii) It is indeterminate whether x^+ is better than y.
>
> In addition, these three claims are jointly compatible with it being determinate that x and y are commensurable.[68]

[65] Espinoza (2008, p. 135).
[66] Espinoza (2008, p. 137).
[67] Espinoza (2008, p. 137).
[68] Rabinowicz (2009, p. 74).

The controversial claims here are premises (i) and (iii). Chang explicitly rejects (i) and (iii) and offers two arguments why the cases in the small-improvement argument does not depend on indeterminacy. To avoid running a straw-man argument Rabinowicz needs to explain why it can plausibly be held that it is indeterminate whether x^+ is better than y. His argument for this is the old epistemic objection: How can we be sure that our judgements in small-improvement cases are correct? However, as mentioned above, Chang has responded to the epistemic objection by offering versions of the small-improvement argument where it seems that we have all the relevant knowledge. Since Rabinowicz does not offer any counter-argument to Chang's answer to the epistemic objection and her two arguments against indeterminacy, his case against the small-improvement argument is incomplete.

Nevertheless, I believe the Rabinowicz diagnosis of the small-improvement argument is on the right track. In 5.4 and 5.5 I will try to answer Chang's arguments against the indeterminacy interpretation of the small-improvement cases.

5.4 *The argument from phenomenology*

Chang calls the cases involved in the small-improvement argument 'superhard cases' and cases where there is a borderline application of a vague predicate, 'borderline cases'.[69] Chang offers two arguments for why the superhard cases cannot all be borderline cases. The first argues that the phenomenology of superhard cases is different from that of borderline cases. Chang writes:

> In borderline cases, insofar as we are willing to judge that the predicate applies, we are also willing to judge that it does not apply. Take for example Herbert, a genuine borderline case of baldness. Insofar as we are willing to call Herbert bald, we are also willing to call him not bald. In superhard cases, things are different. The evidence we have inclines us to the judgment that the one item is not better than the other (and not worse and not equally good). So, for example, our research into the philosophical talents of Aye and Bea incline us to the judgment that Aye is not more philosophically talented than Bea: it seems that this is the case without it also seeming that Aye is more philosophically talented. Thus, in a superhard case, insofar as we are willing to judge that "better than with respect to V" does not apply, we are not also willing to judge that it does apply. In the absence of any explanation for why the phenomenology should be different, there is good reason to think that superhard cases are not cases of vagueness.[70]

Chang seems to argue that in borderline cases we are to some extent willing to say that a certain predicate applies but also to some extent that it does not apply. But in superhard cases

[69] Chang (2002, p. 680).
[70] Chang (2002, p. 682).

one is willing to some extent to judge that a certain predicate does not apply without being willing to any extent to judge that it applies.

However, I fail to see that there is this phenomenal difference in superhard cases, that is, in the examples used in the various versions of the small-improvement argument. Take, for instance, de Sousa's Fairly Virtuous Wife. de Sousa writes that the virtuous wife hesitates between $1,000 and virtue.[71] In this case it seems plausible that the virtuous wife is willing to some extent to judge that the money is better than virtue and also willing to some extent to judge that the money is not better than virtue. This could be part of a plausible explanation of why she hesitates. Similar points can be made for the other versions of the story in the small-improvement argument, like Chang's case with coffee and tea. Furthermore, Chang does not offer any argument that there are any phenomenal difference between standard borderline cases and superhard cases. Hence, a premise of the argument from phenomenology lacks support.

5.5 *The argument from perplexity*

Chang's second argument grants that there is some perplexity in superhard cases over whether one alternative is better than another. The argument from perplexity aims to show that in superhard cases this perplexity is not due to indeterminacy. Chang argues that the perplexity in superhard cases differs from that of borderline cases since it is permissible to resolve the perplexity or indeterminacy by arbitrary stipulation in borderline cases but not in superhard cases. Chang writes the following about borderline cases:

> The resolution of a borderline case lacks what we might call "resolutional remainder": given all the admissible ways in which the case might be resolved, there is no further question as to how resolution should proceed—any admissible resolution will do. We might put the point supervaluationally in this way: given the precisifications of a vague predicate, there is no further question as to how borderline cases should be resolved; they are resolved by arbitrarily opting for one precisification over another.[72]

That is, in borderline cases there are a number of admissible ways to resolve the perplexity and all of them are permitted. Chang contrasts this with the superhard cases:

> In superhard cases, there is resolutional remainder; given a list of admissible ways in which the perplexity might be resolved, there is still a further question as to how the perplexity is to be resolved, for that resolution is not simply given by arbitrarily opting for one admissible resolution over another. Admissible resolutions might be given by weightings of the various respects relevant to the comparison; on one weighting, Mozart

[71] de Sousa (1974, p. 545).
[72] Chang (2002, p. 684).

> is determinately better, while on another, he is determinately worse. It is not appropriate in superhard cases to resolve the perplexity by arbitrarily adopting one weighting rather than another: given the weightings, there is still a further question as to which, if any, weighting one ought to adopt.[73]

Hence, in superhard cases there are, according to Chang, a number of admissible ways to resolve the perplexity but not all of them are permitted. So the difference between borderline cases and superhard cases is supposed to be that in superhard cases there are admissible ways to resolve the perplexity that one ought not to adopt. But if this is the difference between borderline cases and superhard cases, then it seems elusive at best. One wonders, how can a resolution be admissible and, at the same time, be one that one ought not to adopt? Furthermore, if a perplexity concerning whether Mozart is better than Michelangelo, for example, ought to be resolved in the affirmative, then would not a rational agent, rather than being perplexed, judge Mozart to be the better?

Similarly to the argument from phenomenology, the supposed difference between superhard and borderline cases seems elusive. Hence, neither of Chang's arguments against the indeterminacy interpretation of superhard cases convinces.

If one cannot rule out that there may be indeterminacy of 'better' in the superhard cases then the small-improvement argument does not work. If one interprets the negative comparisons in the superhard cases, like 'cup a tastes neither better nor worse than cup b', as $\neg D(aBb)$ and $\neg D(bBa)$ rather than $D(\neg(aBb))$ and $D(\neg(bBa))$, the conflict with axiological completeness disappears. For example, one could interpret Chang's coffee and tea example as follows:

> Suppose you must determine which of a cup of coffee and a cup of tea tastes better to you. The coffee has a full-bodied, sharp, pungent taste, and the tea has a warm, soothing fragrant taste. It is surely possible that you rationally judge that the cup of Sumatra Gold tastes neither *determinately* better nor *determinately* worse than the cup of Pearl Jasmine and that although a slightly more fragrant cup of the Jasmine would taste better than the original, the more fragrant Jasmine would not taste *determinately* better than the cup of coffee.

The trouble is that no plausible transitivity principle would yield that the taste of the cup of Sumatra Gold is determinately neither better, worse, nor equally good as the taste of the cup of Pearl Jasmine. To see this, note that the above story does not rule out that it is indeterminate which of the following combination of value relations holds, where a is the less fragrant cup of the Jasmine, b is the cup of Sumatra Gold, and c is the more fragrant cup of Jasmine:

[73] Chang (2002, p. 685).

(I) $cBa \wedge aBb \wedge cBb$.

(II) $cBa \wedge aEb \wedge cBb$.

(III) $cBa \wedge bBa \wedge cBb$.

(IV) $cBa \wedge bBa \wedge cEb$.

(V) $cBa \wedge bBa \wedge bBc$.

Perhaps (III) could be ruled out as unlikely if the improvement of c over a is sufficiently small. Still, neither of the remaining combinations violates transitivity or, for that matter, axiological completeness.

5.6 *An analysis of parity*

If we accept, as I think we should, that the superhard cases are due to indeterminacy, we do not have to deny that the alternatives involved are on a par. Indeterminacy does not rule out the possibility of parity—indeed it provides a way to analyse parity.

> x is *axiologically on a par* with y if, and only if, it is not determinate that x and y are not equally good.
>
> An agent holds x as *preferentially on a par* with y if, and only if, it is not determinate that the agent is not indifferent between x and y.

The possibility of parity on this analysis, however, does not conflict with the view that if two objects are comparable then they are either better, worse, or equally good. Nevertheless, this does not rule out that there are value (preference) relations other than incomparability that hold when none of the traditional value (preference) relations holds. We will call such relations non-traditional value (preference) relations. In the following two sections I will argue that there may be alternatives between which neither better, worse, nor equally good holds and that there may be non-traditional value relations that hold between these alternatives and, moreover, that there may be non-traditional preference relations that hold between alternatives when neither preference in either direction nor indifference holds. However, since the small-improvement argument does not work once one allows for indeterminate value and preference relations, we need a new argument to counter the view that completeness holds. Furthermore, since parity in the end does not seem to hold when it is false that the traditional relations hold, we need some new value and preference relations in order to establish that there are any non-traditional value and preference. Such a new argument and such new preference and value relations are introduced in Section 7 and Essay IV.

6. Fitting-attitude analyses and value-preference symmetry

This section will critically examine two recent frameworks for value relations, one due to Joshua Gert and one to Wlodek Rabinowicz.[74] I will argue that either their frameworks are inadequate or they lack support. A motivation for both of these frameworks is to make room for additional value relations that hold when none of *better*, *worse*, and *equally good* does.

According to the buck-passing account proposed by Scanlon and others, an object belongs to a certain axiological category if it has properties that provide reasons to take up a certain attitude toward it, or to act in certain ways in regard to it.[75] This approach is not new, however, it follows a long tradition that goes back at least to Brentano.[76] Accounts like this are also called fitting-attitude accounts of value since they analyse the value of an object in terms of what attitude is fitting to have towards the object. On this approach value relations between objects are determined by what preference relation is fitting to have toward the objects. Both Gert and Rabinowicz follow this approach and they analyse value relations in terms of some kind of rationally permissible preferential attitudes.

6.1 *Gert's and Rabinowicz's frameworks*

Gert holds that for some alternatives a rational agent is rationally permitted to have a preference for the alternative with a number of different strengths. These strengths of preference could be thought of as a measure of how much the agent wants the alternative. Gert analyses value relations in terms of what ranges of strengths of preferences are rationally permissible. Better is analysed as follows:

> Range Rule: One item is better than another in a certain respect if the lower bound of the range of the strengths of its relevant rationally permissible preferences is higher than the upper bound of the other's; otherwise the items are not traditionally comparable.[77]

One counter-intuitive implication of Gert's Range Rule is that two items are 'traditionally comparable' if, and only if, one of them is better than the other. If two items are equally good they would be traditionally comparable in the sense that one of the traditional value relations held between them. Gert anticipates this worry and suggests that one modifies the Range Rule to allow for the following analysis of equally good:

> [W]e can easily modify the rule to allow equality in the following circumstances: when two items each have the same unique rationally required strength of preference.[78]

[74] Gert (2004), Rabinowicz (2008).
[75] Scanlon (1998, pp. 95–100).
[76] For an account of the buck-passing tradition see Rabinowicz and Rønnow-Rasmussen (2004).
[77] Gert (2004, p. 505).
[78] Gert (2004, p. 506).

This takes care of the traditional value relations. Then Gert proposes that parity holds between items *A* and *B* in the following kind of case:

> That *A* and *B* have exactly the same range. Thus, for any third item, *C*, the rational status of choosing *A* over *C*, or vice versa, will always be the same as the rational status of choosing *B* over *C*, or vice versa. This case might plausibly be called 'parity'.[79]

As pointed out by Rabinowicz, this is a strange account of parity since it cannot hold in cases like those employed in the small-improvement argument.[80] A cup of tea and a cup of coffee cannot be on a par in Gert's framework unless every improved cup of coffee with better taste than the first cup is not also better than the cup of tea. But this is ruled out in the small-improvement argument since one premise states that the improved cup of coffee is not better than the cup of tea.

Rabinowicz takes a similar but slightly more complex approach. Rather than permissible strengths of preference he analyses value relations in terms of permissible preference orderings. Let *K* be the set of all permissible preference orderings. Rabinowicz makes a number of substantial assumptions about *K*:

> What we can assume, however, is that all the orderings in the 'permissible' class *K* are at least *partial* in the following sense: in every such permissible ordering, (i) preference is a strict partial order, i.e., an asymmetric and transitive relation, (ii) equi-preference (= indifference) is an equivalence relation, i.e., it is transitive, symmetric and reflexive, and (iii) for all items *x* and *y*, if *x* and *y* are equi-preferred, then any item preferred/dispreferred to one of them is respectively preferred/dispreferred to the other.[81]

Given *K*, Rabinowicz defines better, equally good, parity, and incomparability as follows:

(B) *x* is *better* than *y* if and only if *x* is preferred to *y* in every ordering in *K*.[82]

(E) Two items are *equally good* if and only if they are equi-preferred in every ordering in *K*.[83]

(P) *x* and *y* are *on a par* if and only if *K* contains two orderings such that *x* is preferred to *y* in one ordering and *y* is preferred to *x* in the other.[84]

(I) *x* and *y* are *incomparable* if and only if every ordering in *K* contains a gap with regard to *x* and *y*, i.e., neither of these items is preferred to the other, nor are they equi-preferred.[85]

[79] Gert (2004, p. 506).
[80] Rabinowicz (2008, p. 30).
[81] Rabinowicz (2008, p. 37).
[82] Rabinowicz (2008, p. 38).
[83] Rabinowicz (2008, p. 38).
[84] Rabinowicz (2008, p. 39).
[85] Rabinowicz (2008, p. 39).

The transitivity of betterness is arguably a conceptual truth. But as Rabinowicz notes, in his framework, the transitivity of betterness depends on the transitivity of preference which seems to be less firmly established.[86] However, the problem is worsened since Rabinowicz also needs to allow for preferential gaps. As is argued in Essay III, the money-pump argument, which is the standard argument for the transitivity of preference, does not work if one allows for preferential gaps. Thus, on Rabinowicz model, the transitivity of betterness depends on the transitivity of preference which he cannot support in the standard way with the money-pump argument.

6.2 *Value-preference symmetry*

Gert's and Rabinowicz's frameworks both have room for even more value relations than those mentioned above. However, the relations mentioned above are sufficient in order to see that in the frameworks of Gert and Rabinowicz there are value relations that lack a corresponding preference relation. This is so because they both have the value relation parity in addition to better, worse, and equally good but they only have traditional preference relations (and, in Rabinowicz's case, preferential gaps). Hence, they violate the following plausible principle:

> *Value-preference symmetry*
> For every value relation there is a corresponding preference relation and for every preference relation there is a corresponding value relation.

For example, each of the traditional dyadic value relations has a corresponding dyadic preference relation that plays the same role preferentially as the value relation does axiologically:

x better than y	preference for x over y
x worse than y	preference for y over x
x equally good as y	indifference between x and y

Similarly, any value relation that can be defined in terms of the traditional value relations like, for example, *weakly better* has a corresponding preference relation defined correspondingly from the traditional preference relations, in this case, *weakly preferred to*.

Gert argues against such a close connection between values and preferences. He claims that we rarely take our preferences to be uniquely rational.[87] However, this claim at most casts doubt on a principle like the following:

(U) For every preference relation R, there is a value relation V such that xRy is rationally permissible if, and only if, xVy.

[86] Rabinowicz (2008, p. 40).
[87] Gert (2004, p. 494).

But value-preference symmetry does not imply (U). Value-preference symmetry does not say anything about what is rationally permissible, it only claims that there is a one-to-one correspondence between value and preference relations. So Gert's claim does not affect value-preference symmetry directly but in combination with certain types of fitting-attitude analyses.

To make a convincing case against value-preference symmetry one needs a cogent argument for a value relation for which there does not exist a cogent parallel argument for a corresponding preference relation, or vice versa. The trouble for Gert and Rabinowicz is that the most prominent case for a value relation that holds when neither of the traditional value relations holds, the combination of the small-improvement argument and Chang's chaining argument, seems equally applicable for preference relations. At least Chang takes her case to be successful also for preference relations. She writes:

> Perhaps the most striking, the possibility of parity shows the basic assumption of standard decision and rational choice theory to be mistaken: preferring X to Y, preferring Y to X, and being indifferent between them do not span the conceptual space of choice attitudes one can have toward alternatives. Put another way, the "partial orderings" sometimes favoured by such theories will underdescribe the range of choice attitudes a rational agent can have toward alternatives.[88]

As explained in Section 5, I do not think that the small-improvement argument is successful. But my arguments that the small-improvement argument fails due to indeterminacy, seem to work equally well against the axiological as against the preferential versions of the argument. Furthermore, if there is some flaw in my argument then this flaw will probably affect my argument equally in its preferential and axiological renditions. Likewise, Chang's chaining argument does not seem more plausible in an axiological rendition than in an preferential one. The point of the chaining argument is to show that for some pair of options between which none of the traditional relations hold, the options are still comparable. Chang refers to an intuitive notion of comparability that is distinct from that one of the traditional relations holds. Chang offers a nice argument for this intuitive notion based on that there seems be substantial disagreement between a 'dichotomist' and a 'trichotomist'. The dichotomist takes two alternatives to be comparable if, and only if, one of the alternatives is better than the other. The trichotomist takes two alternatives to be comparable if, and only if, one is better than the other or they are equally good.

> If each is offering a merely stipulative definition of the term 'comparability', there is no genuine disagreement between them; each uses the term 'comparability' as she pleases. But intuitively, there is a real issue between them. According to a perfectly intuitive

[88] Chang (2002, p. 666).

> notion of comparability, the dichotomist has made a mistake: she has overlooked the possibility of a third relation, "equally good," that might hold when "better than" and "worse than" do not.[89]

The chaining argument runs as follows: [90]

The chaining argument

(C1) If x and y are comparable and the respects relevant to the comparison between them can be balanced against one another and z is like y except for a small unidimensional change, then x and z are comparable.

(C2) There exist three options x, y, and z such that x and y are comparable and none of the traditional relations holds between x and z and there is a continuum of small unidimensional changes connecting y with z.

(C3) There exist two options x and y such that x and y are comparable and none of the traditional relations holds between x and y.

Chang illustrates the chaining argument with an example with Mozart and Michelangelo who might be thought to be neither more or less creative than each other nor equally as creative:

> According to this principle [(C1)], if Mozart is comparable with Talentlessi, then he is also comparable with Talentlessi+, for the difference between Talentlessi and Talentlessi+ is a small unidimensional one, and by hypothesis, such a difference can't trigger incomparability between evaluatively very different items where before they were comparable. And if Mozart is comparable with Talentlessi+, then applying the principle anew, it follows that he is comparable with Talentlessi++, and so on. Comparability with Mozart is preserved through the continuum of small unidimensional differences, and thus we arrive at the conclusion that Mozart is comparable with Michelangelo. By hypothesis, Mozart is not more or less creative than Michelangelo, and nor are the two equally creative. And yet it seems that they are nevertheless comparable.[91]

A common objection to the chaining argument is directed towards (C1), which Chang calls the small-unidimensional-difference principle. This principle conflicts with a version of the Pareto rule which says:

---

[89] Chang (2002, pp. 662–663).
[90] Chang (2002, pp. 673–675).
[91] Chang (2002, p. 674).

> [U]nless we are equally well-off in each of two states of affairs, one state is better than another if at least one of us is better off than we would be in the other state and none of us is worse off, otherwise the states are incomparable.[92]

For example, consider a state *a* of two individuals whose respective well-being is given by the ordered pair (2,2) and a state *b* of the same individuals at (1,2). The Pareto rule yields that *a* is better than *b* and, thus, comparable. A third state *c* with the individuals at (2,1) is like *a* except for a small unidimensional change. However, the Pareto rule yields that *b* and *c* are incomparable. Thus, we have a counterexample to (C1). To avoid these kinds of counterexamples to the axiological version of the chaining argument one might limit the scope of (C1). This is also Chang's move, which in turn raises worries about ad hocness. The same objection can also be mooted against a preferential version of the chaining argument based on similarly troublesome trade-offs.[93]

This kind of value-preference symmetry is a problem for Gert and Rabinowicz since if the case for a non-traditional value relation that may hold when none of the traditional value relations holds is cogent if, and only if, a parallel case for a corresponding non-traditional preference relation is cogent, then *either* their extended frameworks are unmotivated since there is no reason to think there are non-traditional value relations *or* their frameworks are inadequate since there are non-traditional preference relations and their frameworks do not allow for that.

6.3 *A more straightforward fitting-attitude analysis*

If we accept the value-preference symmetry, as I think we should, and still want to analyse some non-traditional value relations there is no need for complex frameworks like those of Gert and Rabinowicz. In what follows, 'fitting' will be taken to be like 'requirement' in its normative strength rather than like 'permission'. That is, if xPy is fitting then $\neg(xPy)$ cannot also be fitting.

> x is *better* than y if, and only if, it is fitting to prefer x to y.
>
> x and y are *equally good* if, and only if, it is fitting to be indifferent between x and y.
>
> x and y are *axiologically on a par* if, and only if, it is fitting to hold x and y to be preferentially on a par.

[92] Chang (2002, p. 676).

[93] Even if this problem was unique to the axiological version, this would not help Gert and Rabinowicz. This is because they need a cogent case for axiological parity that, *mutatis mutandis*, does not also support preferential parity; not the other way around.

> x and y are *(axiologically) incomparable* if, and only if, it is fitting to have a preferential gap between x and y.

This approach does not imply (U) as long as 'fitting' is understood in some other way than 'rationally required'. It could be a moral concept or perhaps a *sui generis*. This account of axiological parity is an alternative to the one proposed in Section 5.6.

7. Some new preference and value relations

In the previous section we saw that while Gert's and Rabinowicz's frameworks had room for some non-traditional value relations, they did not have any room for the corresponding non-traditional preference relations. In this section I will present a new framework for preference and value relations that have room for both non-traditional preference relations and their corresponding value relations. I will also explore a new argument for the possibility of non-traditional preference and value relations. Previous attempts to establish the possibility of non-traditional value relations have been based on the small-improvement argument. But as argued in Section 5, the small-improvement argument does not work once one allows for indeterminate preference and value relations. Furthermore, in order not to violate value-preference symmetry—see Section 6—one must also explain how there can be non-traditional preference relations if there are non-traditional value relations. Hence, we will begin by exploring the possibility of non-traditional preference relations.

Preference relations have traditionally been conceived behaviouristically by way of a choice function $C(\cdot)$. The $C(\cdot)$ function is usually given a purely dispositional interpretation.[94]

> *Dispositions*
> $C(A) \Leftrightarrow_{def}$ the set of dispositions of the agent to choose from the set of options A.

The traditional way of analysing preference relations can be summarized as follows:

$$\begin{array}{l|l} C(\{x, y\}) = \{x, y\} & xIy \\ C(\{x, y\}) = \{x\} & xPy \\ C(\{x, y\}) = \{y\} & yPx \end{array}$$

However, although standard, the dispositional account of preferences is not without critics. James M. Joyce argues that the dispositional account has some major drawbacks.[95] According to Joyce the dispositional interpretation needs to be abandoned for an analysis based on desires.

[94] See e.g. Arrow (1951, p. 16), Savage (1954, p. 17), Sen (1970, Ch. 1*).
[95] Joyce (1999, pp. 19–22).

Desires
$C(A) \Leftrightarrow_{def}$ the set of options in A that are at least as desirable as every option in A.

Nevertheless, irrespective of whether one opts for the desire or the dispositional approach, the traditional way of analysing preference does not have room for any non-traditional relations, since preference in either direction and indifference exhaust the conceptual possibilities. If one also takes $C(\{x, y\}) = \{\varnothing\}$ to be a possibility then one can allow for the absence of any preferential attitude. Rabinowicz takes the absence of any choice disposition to be a preferential gap.[96]

Essay IV develops a new richer framework for preference relations. Rather than mere choices between alternatives, the new framework analyses preference relations in terms of swaps between alternatives. First we need some new notation:

Swap
$x \rightsquigarrow y \Leftrightarrow_{def}$ swap alternative x for alternative y.

Keep
$\circlearrowright x \Leftrightarrow_{def} x \rightsquigarrow x$.

The idea behind the new swap framework is the following: Instead of defining preference relations in terms of hypothetical choices between the compared alternatives, they are defined in terms of hypothetical choices between keeps and swaps between the compared alternatives. In the swap framework preference relations between alternatives x and y are analysed by the combinations of dispositions of desires, $C(\{x \rightsquigarrow y, \circlearrowright x\})$ and $C(\{y \rightsquigarrow x, \circlearrowright y\})$. Thus, $C(\cdot)$ is applied to two different sets of alternatives. For each of these $C(\cdot)$ may return three different subsets. This yields 3^2 combinations rather than just 3, as mapped out in the following table:

		$C(\{y \rightsquigarrow x, \circlearrowright y\}) =$		
		$\{y \rightsquigarrow x, \circlearrowright y\}$	$\{y \rightsquigarrow x\}$	$\{\circlearrowright y\}$
$C(\{x \rightsquigarrow y, \circlearrowright x\}) =$	$\{x \rightsquigarrow y, \circlearrowright x\}$	xIy		
	$\{x \rightsquigarrow y\}$			yPx
	$\{\circlearrowright x\}$		xPy	

Indifference between x and y is here analysed as the combination of $C(\{x \rightsquigarrow y, \circlearrowright x\}) = \{x \rightsquigarrow y, \circlearrowright x\}$ and $C(\{y \rightsquigarrow x, \circlearrowright y\}) = \{y \rightsquigarrow x, \circlearrowright y\}$. The idea is that if you are indifferent between two alternatives, you do not have any more desire to keep one of them than to swap for the other and vice versa. We assume that there is no transition cost for swapping in these hypothetical choices. We do not want to rule out that you may be indifferent between, for

[96] Rabinowicz (2008, p. 26).

example, some cucumber sandwiches and £2 in ready money and still desire to keep the money over swapping it for the sandwiches if the mere transfer was toilsome.

A preference for x over y is analysed as the combination of $C(\{x \rightharpoonup y, \circlearrowleft x\}) = \{\circlearrowleft x\}$ and $C(\{y \rightharpoonup x, \circlearrowleft y\}) = \{y \rightharpoonup x\}$. If you prefer an alternative over another, you should presumably desire to keep the preferred alternative over swapping to the other one, and further you should desire to swap to preferred alternative over keeping the other.

This accounts for the traditional preference relations. But as seen in the table above, preference in either direction, and indifference do not exhaust the possibilities in the swap framework. I explore these new possible preference relations in Essay IV, defined as follows: [97]

Incommensurateness
$x\#y \Leftrightarrow_{def} C(\{x \rightharpoonup y, \circlearrowleft x\}) = \{\circlearrowleft x\}$ and
$C(\{y \rightharpoonup x, \circlearrowleft y\}) = \{\circlearrowleft y\}$.

Semi-incommensurateness
$xOy \Leftrightarrow_{def} C(\{x \rightharpoonup y, \circlearrowleft x\}) = \{\circlearrowleft x\}$ and
$C(\{y \rightharpoonup x, \circlearrowleft y\}) = \{y \rightharpoonup x, \circlearrowleft y\}$.

Instability
$x*y \Leftrightarrow_{def} C(\{x \rightharpoonup y, \circlearrowleft x\}) = \{x \rightharpoonup y\}$ and
$C(\{y \rightharpoonup x, \circlearrowleft y\}) = \{y \rightharpoonup x\}$.

Semi-instability
$xSy \Leftrightarrow_{def} C(\{x \rightharpoonup y, \circlearrowleft x\}) = \{x \rightharpoonup y, \circlearrowleft x\}$ and
$C(\{y \rightharpoonup x, \circlearrowleft y\}) = \{y \rightharpoonup x\}$.

These new preference relations in addition to the traditional relations exhaust the range of possible preference relations in the swap framework, as can be seen in the following table where all relations have been mapped out:

		$C(\{y \rightharpoonup x, \circlearrowleft y\}) =$		
		$\{y \rightharpoonup x, \circlearrowleft y\}$	$\{y \rightharpoonup x\}$	$\{\circlearrowleft y\}$
$C(\{x \rightharpoonup y, \circlearrowleft x\}) =$	$\{x \rightharpoonup y, \circlearrowleft x\}$	xIy	ySx	yOx
	$\{x \rightharpoonup y\}$	xSy	$x*y$	yPx
	$\{\circlearrowleft x\}$	xOy	xPy	$x\#y$

Here it might help to consider two closely related objections. The swap framework has one drawback in comparison with the traditional framework; it is more complex since it makes use of desires or dispositions over two alternative sets, whereas the traditional framework

[97] Gustafsson (forthcoming-a).

only uses one. Furthermore, what stops us from generating even more preference relations by taking into account the desires or dispositions for even more sets of alternatives? In regard to the first worry I think the swap framework makes up for its additional complexity since, as I argue in Essay IV, it can account for some plausible preferential attitudes that the traditional framework cannot. And in response to the second worry, I think any additional complexity would be unwarranted unless one could account for some plausible preferential attitude that the swap framework cannot account for. We want, echoing Einstein's Razor, a framework that is as simple as possible, but not simpler. If no further value or preference should turn up, that the swap framework cannot account for, no additional complexity is motivated.

The new non-traditional preference relations of the swap framework would be of limited interest if no agent would be rationally permitted to hold them. Nevertheless, I do not want to rule out that there may be more preference relations than those that may be held by rational agents. In the remainder of this section I will present a condensed version of what I believe to be a novel argument against the view that it is rationally required that the weak preference relation is complete. For an examination of these new preference relations and some reasons to believe they might be rationally held, see Essay IV.

The swap argument

(S1) For all x and y and all rational agents P, if P weakly prefers x to y then P holds $y \rightsquigarrow x \in C(\{y \rightsquigarrow x, \circlearrowleft y, \})$.

(S2) For some x and y and some rational agent P, P neither holds $y \rightsquigarrow x \in C(\{y \rightsquigarrow x, \circlearrowleft y\})$ nor $x \rightsquigarrow y \in C(\{\circlearrowleft x, x \rightsquigarrow y\})$.

(S3) It is not rationally required that weak preference is complete.

Premise (S1) states that in a choice between swapping y for x and keeping y, if you weakly prefer x to y then swapping y for x is at least as desirable as keeping y (or one of your dispositions is to swap y for x). Assuming further that the swap in the hypothetical choice involves no transition costs, this premise seems very plausible.

Premise (S2) states that for some pair of objects some rational agent neither desires to swap one for the other nor vice versa (or the agent is neither disposed to swap one for the other nor vice versa). It is made plausible by reference to the standard cases of incommensurable values. Consider, for example, the following two alternatives: [98]

a = \$1,000.

b = friendship.

[98] See e.g. Raz (1988, p. 337).

To most people, I think, friendship is not something to be bought and sold. It seems plausible then that some rational agents would neither be disposed to swap *a* for *b* in a choice between keeping *a* and swapping *a* for *b* nor be disposed to swap *b* for *a* in a choice between keeping *b* and swapping *b* for *a*. It seems plausible also that some rational agents would neither find swapping *a* for *b* at least as desirable as keeping *a* nor would they find swapping *b* for *a* at least as desirable as keeping *b*.

There may also be some empirical support in favour of (S2). In their study of the endowment effect, Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler found that the lowest price at which their test subjects were willing to sell a cup was higher than the highest price at which they were willing to buy the cup. Seemingly rational test subjects were willing to pay up to $2 for a coffee cup while willing to accept no less than $5 for the cup.[99] Thus, the subjects did not desire to swap the following alternatives, in either direction.

c = coffee cup.
d = $4.

The experiments were set up to rule out possible transaction costs.[100] Given that we take these subjects to be rational, the experiments suggest that there are pairs of alternatives such that rational agents may not desire a swap in either direction from one of the alternatives to the other. Thus, they then offer some support for (S2).

Given (S1) and (S2) the conclusion (S3) follows, granted that the set of rational requirements is consistent (this means that there are no dilemmas regarding rational requirements, that is, it is never the case that you are rationally required to *p* and rationally required to not *p*).

Note that this argument differs from the small-improvement argument.[101] Notably, it does not make use of any small improvements. It was the use of small-improvements that made Chang's superhard cases hard to distinguish from borderline cases of a vague predicate. This makes the swap argument less vulnerable to the objection from indeterminacy than the small-improvement argument.

7.1 *Some new value relations*

Given the new non-traditional preference relations in the swap framework, we can provide a fitting-attitude analysis for some new non-traditional value relations. And unlike Gert's and Rabininowicz's frameworks, we do not need to reject value-preference symmetry.

[99] Kahneman et al. (1990b, p. 1332).
[100] Kahneman et al. (1990b, p. 1335).
[101] See § 4 and Essay III.

Axiological incommensurateness
x is axiologically incommensurable with y if, and only if, $x\#y$ is fitting.

Thus, for example, a friendship is axiologically incommensurate with a sum of money if it is fitting to desire to keep the money over swapping it for the friendship and to desire to keep the friendship over swapping it for the money. This fits well with Raz's examples of incommensurateness, of which friendship and money is one.[102] Another of his examples concerns the alternatives having children and having money:

> For such parents, having children and having money cannot be compared in value. Moreover, they will be indignant at the suggestion that such a comparison is possible. Finally, they will refuse to contemplate even the possibility of such an exchange. […] [M]y example concerns […] those for whom it is equally unacceptable to buy a child as to sell one. For them it is not the case that having a child is worth more than any sum of money. If it were then they would not object to buying children when they want them.[103]

Raz, however, takes incommensurateness to be a case of incomparability. Nevertheless, incommensureteness on my analysis is a positive value relation; it is not defined merely as the denial of other value relations.[104]

Here one might object that incommensurateness on my analysis does not seem to hold between the alternatives in many of the examples employed in the small-improvement argument. For example, in Chang's example with tea and coffee, it does not seem unfitting to swap the tea for the coffee nor vice versa. But in such cases, like the tea and coffee example, I think the alternatives are not incommensurate but on a par according to my analysis in 5.6.

Similarly, we can analyse further new dyadic value relations as holding between two objects when the other preference relations in the above framework are fitting between the objects. When two objects x and y are such that it is fitting to desire to swap x for y, or to keep x or y but not to swap y for x when given these alternatives, then the following value relation holds between the objects:

Axiological semi-incommensurateness
x is axiologically semi-incommensurable with y if, and only if, xOy is fitting.

Finally the two more exotic preference relations give us two equally exotic value relations.

Axiological instability
x is axiologically unstable to y if, and only if, $x*y$ is fitting.

[102] See e.g. Raz (1988, pp. 337–338).
[103] Raz (1988, pp. 346–347).
[104] Cf. Chang (2002, p. 663).

Axiological semi-instability
x is axiologically semi-unstable to y if, and only if, xSy is fitting.

This completes our survey of dyadic value relations in terms of fitting swapping attitudes. All the value relations that are analysable in this framework can be mapped out as follows:

Value relation	Fitting attitude
x better than y	xPy
x worse than y	yPx
x equally good as y	xIy
x is axiologically incommensurate with y	$x\#y$
x is axiologically semi-incommensurate with y	xOy
y is axiologically semi-incommensurate with x	yOx
x is axiologically unstable to y	$x*y$
x is axiologically semi-unstable to y	xSy
y is axiologically semi-unstable to x	ySx

8. Preferences and freedom of choice

In the previous sections we have examined what one is rationally required to choose and rational requirements on the structure of our preferences. In addition to rational requirements like those we have discussed so far, there is another restriction on our choices due to what options are available to us. These limitations in what options are available often force us to choose something other than our most preferred alternative among all possible options. Some sets of options restrict our choices in this way less than others. These sets offer us more freedom of choice. In recent years there has been a search for an adequate measure of freedom of choice. Essay V argues that there is a connection between the amount of freedom of choice a set offers and how well it is expected to satisfy an agent with a certain kind of unknown preferences.

Much of the recent debate on freedom of choice has centred around Pattanaik and Xu's axiomatization of the cardinality rule that ranks sets of alternatives with respect to freedom by their cardinality.[105] The cardinality measure has been criticized for neither taking the degree of similarity between offered options nor preferences into account. One influential proposal that takes preferences into account has been put forward by Kenneth Arrow.[106] His account measures freedom of choice by the expected utility a set offers, given a probability distribution over utility functions. This is a promising approach that gives a plausible account

[105] Pattanaik and Xu (1990).
[106] Arrow (1995).

of the relation between freedom of choice and preference. However, while Arrow's approach takes preferences into account, he still does not take the similarity between the offered alternatives into account. The new measure proposed in Essay V is an attempt to develop Arrow's approach to do just that.[107] The measure tries to capture the idea that a set of options offers more freedom of choice the better it represents the set of all possible options. A set of options would be optimal with respect to freedom of choice if it included every possible option and thus, would let one choose anything.

The main guide for my approach is the following type of intuitions. Consider,

> *The Unpredictable Boss*
> Suppose you are going to prepare a set of alternatives from which your boss will choose one. You know that the boss has a favourite alternative in the set of all possible alternatives, and that he wants to choose an alternative that is as similar as possible to his favourite alternative. You estimate that all possible alternatives have the same probability of being the boss's favourite alternative.
>
> Suppose you have to choose between offering the boss one of two different sets of alternatives, *A* and *B*, and you happen to know, never mind why, that *A* offers more freedom of choice than *B*. If you want to minimize the expected degree of dissimilarity between the boss's favourite alternative and the least dissimilar alternative in the set of alternatives you offer him, which of the sets, *A* and *B*, would you offer him?

In this case I think it is very intuitive that you should take *A*. I argue in Essay V that this intuition suggests the following measure of freedom of choice:

> *The expected-compromise measure*
> Given the domain of all possible alternatives Ω, S offers at least as much freedom of choice as S' if, and only if, the expected degree of dissimilarity between a random alternative from Ω and the least dissimilar alternative in S is at least as low as the expected degree of dissimilarity between a random alternative from Ω and the least dissimilar alternative in S'.

To state a more precise version of the measure we need some new notation. Let Ω denote the set of all possible alternatives in the domain, which we assume to be finite. If for example we are measuring the freedom of choice with respect to breakfast cereals that some stores offer, Ω would be the set of all possible breakfast cereals. Let $d(x, y)$ be a function from $\Omega \times \Omega$ to $\mathbb{R}_+$ that measures the degree of dissimilarity between x and y on a ratio scale.

[107] Gustafsson (2010a).

Finally, let $D(x, U)$ be the minimal dissimilarity between the element x and an element in U, $D(x, U) = \min(\{d(x, y) : y \in U\})$.

> *The unweighted expected-compromise measure*
> Given the domain Ω, the non-empty subset U offers at least as much freedom of choice as the non-empty subset V if, and only if,
> $\sum_{x \in \Omega} D(x, U) \leq \sum_{x \in \Omega} D(x, V)$.

In the remainder of this section I will discuss some objections that were only discussed briefly, if at all, in the essay. For a more elaborate defence of the measure see Essay V, which also extends this measure in order to take the values of the alternatives into account.

8.1 *Being able to choose what one prefers*

A first objection is that freedom of choice consists in being able to choose what one prefers. This conception of freedom of choice is not captured by my measure since the measure does not take into account the actual preferences of the agent. However, the idea that freedom of choice consists in being able to choose what one prefers has some very counter-intuitive consequences.

Suppose that you prefer Tie Guanyin over all other teas and that there are two tea shops, *A* and *B*. *A* offers only Tie Guanyin and *B* offers all kinds of tea except, of course, Tie Guanyin. Now the view that freedom of choice is being able to choose what one prefers yields the counter-intuitive result that *A* offers more freedom of choice than *B*. A store that offers almost every kind of tea surely offers more freedom of choice than a store that only offers one obscure tea. In certain cases you are better off choosing from a set of options that offers little freedom of choice than from a set that offers much.

Another closely related objection is that the reasonableness of the expected-compromise measure requires that agents' preference orderings reflect the similarity ordering. If the agent did not prefer alternatives in order of their similarity to one's favourite alternative, why else would the agent want to minimize the expected degree of dissimilarity to her preferred alternative?[108] However, this objection is based on a misunderstanding—my measure does not require that it is in an agent's interest to always have more freedom of choice or, for that matter, a lower expected degree of dissimilarity to her preferred alternative.

The underlying idea that it is *always* in one's interest to have as much freedom of choice as possible is, as I have argued, wrong.

[108] I thank Erik Carlson and Karin Enflo for this objection.

8.2 *Alternatives as dissimilar as possible*

The next objection is based on the intuition that a set of options offers more freedom of choice than another if its options are more dissimilar to each other. More formally it violates the following condition:

> *Weak dominance in dissimilarity (WDD)*
> If S and S' are sets of options with more than one option and S'' is a subset of S' and $\phi(\cdot)$ is a one-to-one function mapping S onto S'' and
> $\forall x \forall y \in S(d(\phi(x), \phi(y)) \geq d(x, y))$,
> then S' offers at least as much freedom of choice as S.

A similar condition has often been proposed in the literature as a condition on the measurement of diversity.[109] To get a feel for the condition it might help with an example. Consider the following sets:

$A = \{$*vote extreme left*, *vote extreme right*$\}$.
$B = \{$*vote moderate left*, *vote moderate right*$\}$.

Since one can map the pair of options in B to the pair of options A so that the elements of A are more dissimilar than those of B, A offers more freedom of choice than B according to WDD. However, I think this is the wrong answer. For most of the views in the political spectrum there is no similar option in A, thus, most of the political spectrum is unrepresented in the choices offered by A. The situation is better in B. Most possible views in the political spectrum are better represented in B than in A, and thus B offers more freedom of choice with respect to the total political spectrum. The only views better represented in A than in B are the views more similar to the extreme views than the moderate views, and that range is smaller than the range that is better represented in B. Thus, B offers more freedom of choice than A.

For similar reasons I think that the corresponding condition on the measurement of diversity fails. In order to have diversity in for example the animal kingdom, the set of all possible species should be as well represented as possible. This intuition is incompatible with the view that the species in the animal kingdom should be as dissimilar as possible, which follows *mutatis mutandis* from the counter-argument to WDD above.

8.3 *Evenness of dissimilarity between options*

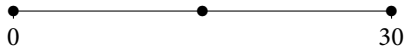
Another objection is based on the idea that an even distribution of the degrees of dissimilarity between different pairs in a set contributes positively to the amount of freedom of choice that

[109] See e.g. Weitzman (1992, p. 392), Pattanaik and Xu (2008, p. 263).

the set offers. And furthermore, the expected-compromise measure does not reflect this.[110] For an example of the idea, suppose that we want to measure the amount of freedom of choice with respect to temperatures between 0 °C and 30 °C. This could be the temperature you could have in your office and we are comparing different thermostats that offer you different temperature options. Consider the following sets:

$A = \{0\ °C, 1\ °C, 30\ °C\}$.

$B = \{0\ °C, 15\ °C, 30\ °C\}$.



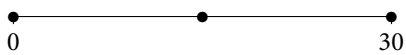
Intuitively, *B* offers more freedom of choice than *A*. One explanation for this could be that the degrees of dissimilarity between the options in the different pairs in the sets are more even in *B* than in *A*. Cases like this might give the evenness-of-dissimilarities-between-options idea some support unless, of course, there is a better explanation of why *B* offers more freedom of choice than *A*. But there are better explanations. An obvious candidate is that a large range of possible alternatives is better represented in *B* than in *A* and just a small range of possible alternatives is better represented in *A* than in *B*.

Furthermore, the evenness-of-dissimilarities-between-options explanation has some serious problems. It conflicts with more intuitive principles. One can show that the evenness-of-dissimilarities-between-options idea conflicts with the following monotonicity condition.

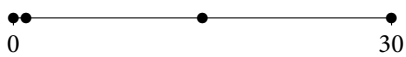
Monotonicity
For all sets *S* and all options *x*, $S \cup \{x\}$ offers at least as much freedom of choice as *S*.

Monotonicity is the least controversial principle in the literature on how to measure freedom of choice.[111] This is so for fairly obvious reasons. How can the freedom of choice offered be lessened by the addition of a new alternative with all previous alternatives retained? The trouble is that monotonicity severely limits the positive contribution that evenness can make to the freedom of choice offered by a set. Consider the following sets:

$B = \{0\ °C, 15\ °C, 30\ °C\}$.



$C = \{0\ °C, 1\ °C, 15\ °C, 30\ °C\}$.



Set *C* does not seem to be much better than *B* with respect to freedom of choice. The only difference is the addition of an alternative that is very similar to an already existing one and this should contribute only minimally to the amount of freedom of choice offered. However,

[110] This objection is due to Karin Enflo.
[111] See e.g. Pattanaik and Xu (1990, p. 386).

with respect to evenness of dissimilarities between options C is much worse than B. The problem is that monotonicity demands that C offers at least as much freedom of choice as B. If evenness of dissimilarities between options had been an important feature for freedom of choice, then the large loss of evenness from B to C would have outweighed the very minor respects in which C could be thought to improve on the freedom of choice offered by B. Thus, if monotonicity holds, then evenness of dissimilarities of options cannot contribute much to the amount of freedom of choice that a set offers. Furthermore, the problem would be sharpened if the option 1 °C was replaced by one even more similar to 0 °C. The evenness of this new set would be even smaller and the respects in which it improves on the freedom of choice offered by B would be even less. Thus, the evenness of dissimilarities of options does not seem to contribute at all to the freedom of choice a set offers.

A possible reply in defence of the evenness-of-dissimilarities-between-options explanation is that it only applies under special circumstances like for example when the sets have the same number of options or when the sum of all degrees of dissimilarities between all pairs is equal in the sets. However, such a reply would be extremely ad hoc if one does not also explain why evenness only contributes to freedom of choice under these circumstances. But the prospects for such an explanation looks dim. Thus, the evenness-of-dissimilarities-between-options idea lacks support.

As an example of this last type of approach we will examine a proposal by Karin Enflo. Enflo has proposed a family of measures of diversity and freedom of choice that are supposed to satisfy monotonicity and to favour sets with evenly distributed options when comparing sets with the same number of options. The measures are stated as follows:

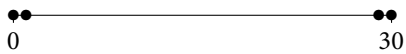
> *The Ratio root measures*:
> $D(A^*) = D(A,d) = \beta\frac{1}{n-1}\sum_{i=1,i\neq j}^{n}\sum_{j=1}^{n}(d(x_i,x_j))^r$ with $\beta > 0$, $\frac{1}{2} \leq r < 1$ and $n \geq 2$,
> where D is a function from the family of finite metric spaces to $\mathbf{R}$, A^* is a metric space, A is a set, d is a function from $A \times A$ to $\mathbf{R}$, $n = |A|$ and $x_1 \ldots x_n$ are the elements of A.[112]

A major worry about the ratio root measures is that they are, like other measures that depend on the sum of all distances between options, unreasonably sensitive to the number of alternatives.

Consider the following sets:

$B = \{0\ °C, 15\ °C, 30\ °C\}$.

$E = \{0\ °C, 1\ °C, 29\ °C, 30\ °C\}$.



[112] Enflo (2011).

B seems to offer more diversity and more freedom of choice than E. B offers in addition to a very cold and a very hot option, a moderately warm option at 15 °C whereas E only offers very cold and very hot alternatives. Furthermore, intuitively, the options in B are more evenly distributed than the options in E are. The trouble is that all the ratio root measures rank E over B. This is so with any combination of admissible values of β and r. Choose, for example, $\beta = 1$ and $r = 1/2$. Then we have that $D(B) \approx 13.2$ and $D(E) \approx 15.7$. Moreover, E would be ranked higher than B by all ratio root measures even if the option 1 °C was replaced in E by an option arbitrarily more similar to 0 °C and if the option 29 °C was replaced in E by an option arbitrarily more similar to 30 °C. Hence, the ratio root measures do not adequately measure freedom of choice or for that matter diversity. Note also that the ratio root measures do not seem to favour sets with evenly distributed options in general, since they rank E over B while the options in B are also, in addition to their other advantages, much more evenly distributed than the options in E.

8.4 *No-choice situations*

A prima facie plausible principle that conflicts with the expected-compromise measure is the principle of indifference between no-choice situations: [113]

> *Indifference between no-choice situations (INS)*
> All choice sets with only one alternative offer equally as much freedom of choice.

The intuition behind this principle is that choice sets with only one alternative offer no choice at all, thus they all offer the same amount of freedom of choice.

The strongest argument in the literature against indifference of no-choice situations is due to Peter Jones and Robert Sugden. They argue that INS is incompatible with the conjunction of the following two principles:

> [T]he *Principle of Addition of Significant Options*: [I]f one choice set S contains every option that is contained in another choice set S' and also contains an additional option x, then if x is significant in relation to S, S offers a greater [amount of freedom] of choice than S'.
>
> [T]he *Principle of Addition of Insignificant Options*: [I]f one choice set S contains every option that is contained in another choice set S' and also contains just one additional

[113] Note that the expected-compromise measure can with a slight modification be made compatible with indifference between no-choice situations, see Essay V fn. 22.

> option x, then if x is not significant in relation to S, S offers exactly the same [amount of freedom] of choice as S'.[114]

However, in order for the conclusion to follow one needs a further, albeit very plausible, principle:

> *Coexistence of significant and insignificant options*
> There are choice sets $\{x\}$, $\{y\}$, and $\{x, y\}$ such that x is significant in relation to $\{x, y\}$ and y is insignificant in relation to $\{x, y\}$.

Consider the following example where a = *staying in cell* and b = *being shot*. Since a is significant in relation to $\{a, b\}$, $\{a, b\}$ offers a greater amount of freedom of choice than $\{b\}$, and since b is not significant in relation to $\{a, b\}$, $\{a, b\}$ offers an equal amount of freedom of choice as $\{a\}$. By transitivity it follows that $\{a\}$ offers more freedom of choice than $\{b\}$, and thus, that INS is false.

However, a weak link in the argument is the principle of addition of insignificant options. It seems a bit too strong to claim that the addition of an insignificant option cannot bring about any increase in freedom of choice. Even though *being shot* in the example is an insignificant addition, $\{a, b\}$ might still seem to offer a little bit more freedom of choice than $\{a\}$.

Therefore I suggest that we replace the principle of insignificant options with a more plausible candidate:

> *The weak principle of comparative significance*
> If one choice set S includes the options x and y, and x is more significant in relation to S than is y, and S' consists of all options in S except x and S'' consists of all options in S except y, then S'' offers a greater amount of freedom of choice than S'.

Even though one might find it implausible that there would be no gain in freedom of choice if one, in addition to staying in the cell, was given the option of being shot, one should agree that this gain is less than the gain in freedom of choice if one in addition to being shot was given the option of staying in the cell. It then follows that $\{a\}$ must offer more freedom of choice than $\{b\}$. And since $\{a\}$ and $\{b\}$ are both singleton sets this implies that INS is false.

Another principle concerning no-choice situations that the expected-compromise measure violates is the following:

[114] Jones and Sugden (1982, p. 57). Instead of the amount of freedom of choice, Sugden and Jones's principles concern the value of freedom of choice a choice set offers, which is irrelevant for my purposes. I think the argument works equally well in my modified rendition. The basic idea is the same.

> *Dominance of multi-choice situations (DMS)*
> Any choice set with two or more options offers more freedom of choice than any set with only one option.[115]

The intuition behind DMS seems to be that multi-option choice sets give the agent a choice whereas singleton choice sets give the agent none. Thus, multi-option sets offer more freedom of choice. This principle is open to objections similar to those offered against INS. However, they need a slightly stronger version of the comparative significance principle:

> *The strong principle of comparative significance*
> If one choice set S includes the sets of options U and V, and U is more significant in relation to S than is V, and S' consists of all options in S except those in U and S'' consists of all options in S except those in V, then S'' offers a greater amount of freedom of choice than S'.

In addition to the options a = *staying in cell* and b = *being shot*, we have a third option, c = *being hanged*. Like in the above example it seems plausible that the gain in freedom of choice of adding the options of being hanged and being shot to the option of staying in the cell is less than the gain of adding the option of staying in cell to being hanged and being shot. It then follows that $\{a\}$ offers more freedom of choice than $\{b, c\}$. This in turn implies that DMS is false.

Since the weak and strong principles of comparative significance are not quite self evident, one might still be unwilling to give up INS and DMS. However, as I mention in Essay V, one can easily modify the expected-compromise measure to make it compatible with both INS and DMS: [116]

> *The INS-expected-compromise measure*
> Given the domain Ω, the non-empty subset U offers at least as much freedom of choice as the non-empty subset V if, and only if,
>
> 1. V is a singleton set, *or*
> 2. neither U nor V is a singleton set, and $\sum_{x \in \Omega} w(x)D(x, U) \leq \sum_{x \in \Omega} w(x)D(x, V)$.

This measure behaves exactly like the standard expected-compromise measure except that it ranks all singleton sets lower than any multi-option set and it ranks all singleton sets as equal with respect to offered freedom of choice. One might object that the INS-expected

[115] This principle was suggested by Erik Carlson who takes it to be a conceptual truth.
[116] Gustafsson (2010a, p. 76).

compromise measure is ad hoc since it handles singleton sets differently in order to satisfy INS and DMS. But one can reply that there is a rationale for ranking singleton sets differently: Multi-option sets offer agents a choice whereas singleton sets do not. This can explain the different treatment of singleton sets. One might object that this rationale makes the account circular—whether a set of options offers the agent a choice is something that the account should provide an answer to. But the theory is a theory of freedom of choice, not a theory of when one has a choice. So the rationale is not circular.

Thus, even *if* one accepts the intuitions behind INS and DMS, one does not have to completely give up the approach of the expected-compromise measure.

References

Arrow, Kenneth J. (1951) *Social Choice and Individual Values*, New York: Wiley.

— (1995) 'A Note on Freedom and Flexibility', in Kaushik Basu, Prasanta K. Pattanaik, and Kotaro Suzumura, eds., *Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen*, pp. 7–15, Oxford: Clarendon Press.

Bourget, David and David Chalmers (2009) 'The PhilPapers Surveys', `http://philpapers.org/surveys/`.

Broome, John (1997) 'Is Incommensurability Vagueness?', in Ruth Chang, ed., *Incommensurability, Incomparability, and Practical Reason*, pp. 67–89, Cambridge, MA: Harvard University Press.

— (2004) *Weighing Lives*, Oxford: Oxford University Press.

— (2009) 'Reply to Rabinowicz', *Philosophical Issues* 19 (1): 412–417.

Cantwell, John (2010) 'On an Alleged Counter-Example to Causal Decision Theory', *Synthese* 173 (2): 127–152.

Carlson, Erik (2004) 'Broome's Argument against Value Incomparability', *Utilitas* 16 (02): 220–224.

— (forthcoming) 'The Small-Improvement Argument Rescued', *The Philosophical Quarterly*.

Chang, Ruth (1997) 'Introduction', in Ruth Chang, ed., *Incommensurability, Incomparability, and Practical Reason*, pp. 1–34, Cambridge, MA: Harvard University Press.

— (2002) 'The Possibility of Parity', *Ethics* 112 (4): 659–688.

Davidson, Donald, J. C. C. McKinsey, and Patrick Suppes (1955) 'Outlines of a Formal Theory of Value, I', *Philosophy of Science* 22 (2): 140–160.

de Sousa, Ronald (1974) 'The Good and the True', *Mind* 83 (332): 534–551.

Eells, Ellery (1982) *Rational Decision and Causality*, Cambridge: Cambridge University Press.

Egan, Andy (2007) 'Some Counterexamples to Causal Decision Theory', *The Philosophical Review* 116 (1): 93–114.

Enflo, Karin (2011) 'Measuring Diversity of Choice Sets', unpublished manuscript.

Espinoza, Nicolas (2008) 'The Small Improvement Argument', *Synthese* 165 (1): 127–139.

Fisher, Ronald A. (1957) 'Dangers of Cigarette-Smoking', *British Medical Journal* 2 (5039): 297–298.

Gardner, Martin (1986) *Knotted Doughnuts and Other Mathematical Entertainments*, New York: Freeman.

Gert, Joshua (2004) 'Value and Parity', *Ethics* 114 (3): 492–510.

Gibbard, Allan and William L. Harper (1978) 'Counterfactuals and Two Kinds of Expected Utility', in C. A. Hooker Hooker, J. J. Leach, and E. F. McClennen, eds., *Foundations and Applications of Decision Theory*, vol. I, pp. 125–162, Dordrecht: Reidel.

Gustafsson, Johan E. (2010a) 'Freedom of Choice and Expected Compromise', *Social Choice and Welfare* 35 (1): 65–79.

— (2010b) 'A Money-Pump for Acyclic Intransitive Preferences', *Dialectica* 35 (2): 251–257.

— (forthcoming-a) 'An Extended Framework for Preference Relations', *Economics and Philosophy*.

— (forthcoming-b) 'A Note in Defence in Ratificationism', *Erkenntnis*.

Gustafsson, Johan E. and Nicolas Espinoza (2010) 'Conflicting Reasons in the Small-Improvement Argument', *The Philosophical Quarterly* 60 (241): 754–763.

Hansson, Sven Ove (1993) 'Money-Pumps, Self-Torturers and the Demons of Real Life', *Australasian Journal of Philosophy* 71 (4): 476–485.

Hansson, Sven Ove and Till Grüne-Yanoff (2009) 'Preferences', in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, CSLI, spring 2009 edn., `http://plato.stanford.edu/archives/spr2009/entries/preferences/`.

Harper, William L., Robert Stalnaker, and Glenn Pearce (1981) *Ifs: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht: Reidel.

Jeffrey, Richard C. (1965) *The Logic of Decision*, New York: McGraw-Hill.

— (1983) *The Logic of Decision*, Chicago: University Of Chicago Press, second edn.

Jones, Peter and Robert Sugden (1982) 'Evaluating Choice', *International Review of Law and Economics* 2 (1): 47–65.

Joyce, James M. (1999) *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press.

— (forthcoming) 'Regret and Instability in Causal Decision Theory', *Synthese*.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1990b) 'Experimental Tests of the Endowment Effect and the Coase Theorem', *The Journal of Political Economy* 98 (6): 1325–1348.

Lewis, David (1979) 'Prisoners' Dilemma is a Newcomb Problem', *Philosophy and Public Affairs* 8 (3): 235–240.

— (1981) 'Causal Decision Theory', *Australasian Journal of Philosophy* 59 (1): 5–30.
Luce, R. Duncan (1956) 'Semiorders and a Theory of Utility Discrimination', *Econometrica* 24 (2): 178–191.
McClennen, Edward F. (1990) *Rationality and Dynamic Choice: Foundational Explorations*, Cambridge: Cambridge University Press.
Nozick, Robert (1963) *The Normative Theory of Individual Choice*, Ph.D. thesis, Princeton University.
— (1969) 'Newcomb's Problem and Two Principles of Choice', in Nicholas Rescher, ed., *Essays in Honor of Carl G. Hempel*, pp. 114–146, Dordrecht: Reidel.
Pattanaik, Prasanta K. and Yongsheng Xu (1990) 'On Ranking Opportunity Sets in Terms of Freedom of Choice', *Recherches Economiques de Louvain* 56 (3-4): 383–390.
— (2008) 'Ordinal Distance, Dominance, and the Measurement of Diversity', in Prasanta K. Pattanaik, Koichi Tadenuma, Yongsheng Xu, and Naoki Yoshihara, eds., *Rational Choice and Social Welfare*, pp. 259–269, Berlin: Springer.
Peterson, Martin (2007) 'Parity, Clumpiness and Rational Choice', *Utilitas* 19 (4): 505–513.
Rabinowicz, Wlodek (1989) 'Stable and Retrievable Options', *Philosophy of Science* 56 (4): 624–641.
— (2008) 'Value Relations', *Theoria* 74 (1): 18–49.
— (2009) 'Incommensurability and Vagueness', *Aristotelian Society Supplementary Volume* 83 (1): 71–94.
Rabinowicz, Wlodek and Toni Rønnow-Rasmussen (2004) 'The Strike of the Demon: On Fitting Pro-Attitudes and Value', *Ethics* 114 (3): 391–423.
Raz, Joseph (1988) *The Morality of Freedom*, Oxford: Clarendon Press.
Regan, Donald H. (1988) 'Authority and Value: Reflections on Raz's Morality of Freedom', *Southern California Law Review* 62:995–1095.
Savage, Leonard J. (1951) 'The Theory of Statistical Decision', *Journal of the American Statistical Association* 46 (253): 55–67.
— (1954) *The Foundations of Statistics*, New York: Wiley.
Scanlon, T. M. (1998) *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
Schick, Frederic (1986) 'Dutch Bookies and Money Pumps', *The Journal of Philosophy* 83 (2): 112–119.
Sen, Amartya K. (1970) *Collective Choice and Social Welfare*, San Francisco: Holden-Day.
Sinnott-Armstrong, William (1985) 'Moral Dilemmas and Incomparability', *American Philosophical Quarterly* 22 (4): 321–329.
Weirich, Paul (1986) 'Decisions in Dynamic Settings', in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 438–449, Philosophy of Science Asso-

ciation.
— (1988) 'Hierarchical Maximization of Two Kinds of Expected Utility', *Philosophy of Science* 55 (4): 560–582.
Weitzman, Martin L. (1992) 'On Diversity', *The Quarterly Journal of Economics* 107 (2): 363–405.

ANNOTATED ESSAY SUMMARIES

ESSAY I:
A NOTE IN DEFENCE OF RATIFICATIONISM

Essay I defends ratificationism from a recent attack by Andy Egan. Egan argues that neither evidential nor causal decision theory gives the intuitively right recommendation in the cases *The Smoking Lesion*, *The Psychopath Button*, and *The Three-Option Smoking Lesion*. Furthermore, Egan argues that we cannot avoid these problems by any kind of ratificationism. This essay develops a new version of ratificationism, iterated general ratificationism, that yields the intuitively right recommendations. Thus, the new proposal has advantages over evidential and casual decision theory and standard ratificationist evidential decision theory.

The essay introduces a weakened concept of ratifiability, general ratifiability:

> An act A is *generally ratifiable*$_0$ if, and only if, there is no alternative B such that for every alternative C, the unconditional expected utility of B exceeds the unconditional expected utility of A on the supposition that C is decided upon.
>
> An act A is *generally ratifiable*$_{n+1}$ if, and only if, there is no generally ratifiable$_n$ alternative B such that for every generally ratifiable$_n$ alternative C, the unconditional expected utility of B exceeds the unconditional expected utility of A on the supposition that C is decided upon.
>
> An act A is *iteratively generally ratifiable* if, and only if, for all $k \geq 0$, A is generally ratifiable$_k$.

Given the above definitions of general ratifiability, iterated general ratificationism can be stated as follows:

> It is rational to decide upon an act A if, and only if, A is iteratively generally ratifiable and there is no other iteratively generally ratifiable option with higher conditional expected utility than A.

ESSAY II:
A MONEY-PUMP FOR ACYCLIC INTRANSITIVE PREFERENCES

Essay II develops a new version of the money-pump argument for the claim that rational preferences are transitive. The standard money pump only exploits agents with cyclic strict preferences. In order to pump agents who violate transitivity but without a cycle of strict preferences, one needs to somehow induce such a cycle. Methods for inducing cycles of strict preferences from non-cyclic violations of transitivity have been proposed in the literature, based either on offering the agent small monetary transaction premiums or on multi-dimensional preferences.

This essay argues that the approach with small monetary transaction premiums begs the question since it needs to rely on the transitivity of preference in one crucial step. The multi-dimensional approach does not succeed since the approach only work in special cases and, hence, cannot show that intransitive preferences are irrational in general. To overcome these problems, the essay presents a new approach based on the dominance principle. It is shown that an agent with intransitive preferences either violates the dominance principle or has cyclic preferences over a set of lotteries.

ESSAY III:
CONFLICTING REASONS IN THE SMALL-IMPROVEMENT ARGUMENT

Essay III examines the small-improvement argument. This argument is usually considered the most powerful argument against completeness, namely, the view that for any two alternatives an agent is rationally required either to prefer one of the alternatives to the other or to be indifferent between them. The essay argues that while there might be reasons to believe each of the premises in the standard version of the small-improvement argument, there is a conflict between these reasons. As a result, the reasons do not provide support for believing the conjunction of the premises. Without support for the conjunction of the premises, the standard version of the small-improvement argument against completeness fails.

Moreover, the essay argues that the money-pump argument for the irrationality of intransitive preferences does not work if one allows for the possibility of incomplete preferences. Also, the essay presents and defends a principle for the relation between the reasons to believe the individual conjuncts and their support for belief in the conjunction. This principle states that a collection of reasons to believe the individual conjuncts of a conjunction provides a reason to believe the conjunction only if these reasons are reasons to believe each conjunct under the assumption that the other conjuncts are true.

Note to Essay III

While the argument in the essay is valid for the standard version of the small-improvement argument, some new variants of the small-improvement argument seem to lack any problematic conflict between reasons. See section 4 of the introduction for details. I still hold that these new variants are unconvincing but this is for reasons due to indeterminacy. I make my case for this claim in section 5 of the introduction.

ESSAY IV:
AN EXTENDED FRAMEWORK FOR PREFERENCE RELATIONS

Essay IV models preference relations. In order to account for non-traditional preference relations the essay develops a new, richer framework for preference relations. This new framework provides characterizations of non-traditional preference relations, such as incommensurateness and instability, that may hold when neither preference nor indifference do. The new framework models relations with swaps, which are conceived of as transfers from one alternative state to another. The traditional framework analyses dyadic preference relations in terms of a hypothetical choice between the two compared alternatives. The swap framework extends this approach by analysing dyadic preference relations in terms of two hypothetical choices: the choice between keeping the first of the compared alternatives or swapping it for the second; and the choice between keeping the second alternative or swapping it for the first.

The essay also argues that the swap framework makes possible a new interpretation of the much discussed endowment effect. The endowment effect is the hypothesis that utility of an object is higher for agents once they own it. For some test subjects the maximum amount they are willing to pay for a certain coffee cup is lower than the minimum amount they are willing to sell it for. It seems then that the subjects' preference between the cup and different amounts of money changes when they acquire the cup, which seems strange. On the swap framework the subjects' preferences may have been unaffected by the acquisition of the cup.

ESSAY V:
FREEDOM OF CHOICE AND EXPECTED COMPROMISE

Essay V develops a new measure of freedom of choice based on the proposal that a set offers more freedom of choice than another if, and only if, the expected degree of dissimilarity between a random alternative from the set of possible alternatives and the most similar offered alternative in the set is smaller. Furthermore, a version of this measure is developed that is able to take into account the values of the possible options.

The essay also argues that the expected-compromise measure can be conceived as an extension of Kenneth Arrow's preference based approach to freedom of choice. Furthermore, the essay explores the expected-compromise measure's connection to Claus Nehring and Clemens Puppe's very influential multi-attribute approach. I prove that the expected-compromise measure is equivalent to a version of the multi-attribute approach. That is, with certain constraints on the attribute weights, the multi-attribute approach and the expected-compromise measure rank sets of alternatives equivalently.

Note to Essay V

The figures of the essay did not come out fully as intended in the published version. See Appendix B for improved versions of the figures. Furthermore, on the next to last line of page 70 (by journal numbering), ignore the comma after 'however'.

[Essays removed due to copyright]

APPENDICES

A. Proofs

This appendix contains proofs of observations in the essays that were left out due to space limitations and/or triviality. They are included here for completeness.

Observation 1: *The following three statements cannot all be true:*

(i) *The virtuous-wife preferences are rational.*

(ii) *PI-transitivity is rationally required.*

(iii) *Completeness is rationally required.*[1]

Proof:

(1)	$\neg(aPb) \wedge \neg(bPa) \wedge cPa \wedge \neg(cPb)$	The virtuous-wife preferences
(2)	$\forall x \forall y \forall z((xPy \wedge yIz) \rightarrow xPz)$	*PI*-transitivity
(3)	$\forall x \forall y(xPy \vee yPx \vee xIy)$	Completeness
(4)	$\neg(aPb) \wedge \neg(bPa) \wedge \neg(aIb)$	(1), (2)
(5)	$\neg\forall x \forall y(xPy \vee yPx \vee xIy)$	(4)
(6)	$\perp$	(3), (5)

Since the virtuous-wife preferences, *PI*-transitivity, and completeness cannot all be true, the virtuous wife preferences cannot be rational if *PI*-transitivity and completeness are rationally required, if the set of rational requirements is consistent. ∎

Observation 2: *The weighted expected-compromise measure satisfies strong monotonicity.*[2]

Proof: [3] Since S is a subset to $S \cup \{x\}$ it holds that for each possible alternative in the domain Ω that the degree of dissimilarity to the least dissimilar alternative in $S \cup \{x\}$ cannot be higher than the degree of dissimilarity to the least dissimilar alternative in S. $d(x, x) = 0$

[1] See Essay III, Gustafsson and Espinoza (2010, pp. 756–757). In the essay completeness is called comparability.

[2] See Essay V, Gustafsson (2010, p. 75).

[3] For a proof that the unweighted expected-compromise measure satisfies monotonicity, set all weights to 1.

and x being such that for all $y \in S$, $d(x, y) > 0$ implies that $x \notin S$. Because x is a member of $S \cup \{x\}$ and $d(x, x) = 0$, the degree of dissimilarity from the possible alternative $x \in \Omega$ to the least dissimilar alternative in $S \cup \{x\}$ is 0. Since x is such that for all $y \in S$, $d(x, y) > 0$ the degree of dissimilarity between the possible alternative $x \in \Omega$ and the least dissimilar alternative in S is higher than 0. Therefore there is at least one alternative in Ω with a positive weight between which the degree of dissimilarity is lower to the least dissimilar alternative in $S \cup \{x\}$ than to the least dissimilar alternative in S, as $w(x) > 0$. We have therefore that for all $u \in \Omega$, $w(x) \min(d(u, v) : v \in S \cup \{x\}) \leq w(x) \min(d(u, v) : v \in S)$, and that there exist an alternative $p \in \Omega$ such that $w(x) \min(d(p, v) : v \in S \cup \{x\}) < w(x) \min(d(p, v) : v \in S)$. Thus we have that $\sum_{x \in \Omega} w(x) \min(d(x, z) : z \in S \cup \{x\}) < \sum_{x \in \Omega} w(x) \min(d(x, z) : z \in S)$. Which in conjunction with the expected-compromise measure implies that $S \cup \{x\}$ offers more freedom of choice than S. ∎

B. Alternative figures

The figures in the published version of Essay V did not turn out as I intended. Although barely noticeable at first glance, the scales vary between figures 1–4 and also between figures 5 and 6. This makes the figures potentially misleading as support for the relevant comparisons in the essay. This appendix includes the same figures but with consistent scales between figures 1–4 and between figures 5 and 6.[4]

[4] The original figures occur in Gustafsson (2010); figures 1 and 2 on page 67, figures 3 and 4 on page 72, and figures 5 and 6 on page 75.

Figure 1: The sets A and B.

$A = \{0\,°\text{C}, 1\,°\text{C}\}$.

$B = \{0\,°\text{C}, 30\,°\text{C}\}$.

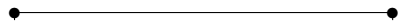
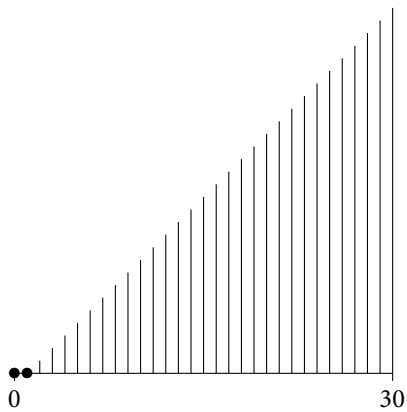


Figure 2: The set C.

$C = \{0\,°\text{C}, 15\,°\text{C}, 30\,°\text{C}\}$.

Figure 3: The dissimilarity with the most similar alternative in A and B for all possible alternatives.

$A = \{0\,°\text{C}, 1\,°\text{C}\}$.



$B = \{0\,°\text{C}, 30\,°\text{C}\}$.

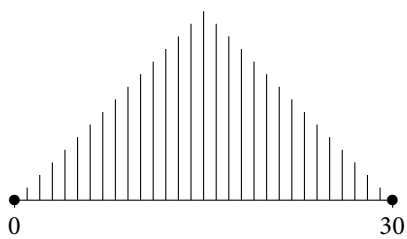


Figure 4: The dissimilarity with the most similar alternative in C for all possible alternatives.

$C = \{0\,°\text{C}, 15\,°\text{C}, 30\,°\text{C}\}$.

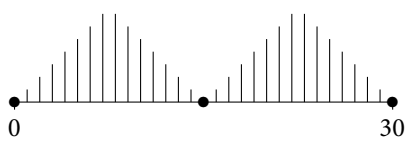


Figure 5: Sets F and G given domain Ω_1.

$F = \{10\,°\text{C}, 20\,°\text{C}\}$.

$G = \{12\,°\text{C}, 18\,°\text{C}\}$.

Figure 6: Sets F and G given domain Ω_2.

$F = \{10\,°\text{C}, 20\,°\text{C}\}$.

$G = \{12\,°\text{C}, 18\,°\text{C}\}$.

References

Gustafsson, Johan E. (2010) 'Freedom of Choice and Expected Compromise', *Social Choice and Welfare* 35 (1): 65–79.

Gustafsson, Johan E. and Nicolas Espinoza (2010) 'Conflicting Reasons in the Small-Improvement Argument', *The Philosophical Quarterly* 60 (241): 754–763.