Should I pretend I'm perfect?

Julia Staffel (University of Southern California)

Introduction

This paper is about the relationship between ideal agents, who obey certain ideal norms perfectly, and non-ideal agents, like humans. Ideal agents are often used in philosophy to illustrate certain ideal norms – norms which humans can never fully comply with because of their limitations and imperfections. We can think of ideal agents as role models in a certain sense whose perfection we try to approximate. However, which form this striving should take is a substantive question. Should we simply try to imitate our ideal counterparts, irrespective of the differences between us and them? Several authors have argued that as far as practical norms are concerned, this is not a good idea. If non-ideal agents try to do what their ideal counterpart would do, the results can be disastrous.

Yet, this question has not received much attention in epistemology, even though epistemology is certainly an area in which formulating ideal norms is an important part of research. A subfield of epistemology that is explicitly concerned with formulating principles of rational belief and belief change is Bayesian epistemology. The principles of Bayesianism are commonly held to constrain the degrees of belief of ideally rational agents. My goal is to consider what happens if a non-ideal agent were to attempt to reason in the same way as an ideal Bayesian agent. If this turns out to produce results that lead the agent further away from complying with the ideal norms, then we can conclude that, just as in the practical cases, it is not advisable for non-ideal agents to imitate ideal agents. However, if imitating ideal agents does not lead the non-ideal agent further away from complying with the ideal norms, then this would be an interesting disparity between the theoretical and the practical realms.

I will first introduce the notion of an ideal agent, and rehearse the arguments against imitating ideal agents in the practical realm. Then I will move on to the epistemological realm by briefly summarizing the Bayesian view of ideal norms on degrees of belief. In the third section, I will lay out my strategy for determining what happens if a non-ideal reasoner tries to follow the rules of reasoning that are designed for

ideal reasoners. In order to be able to assess reasoning strategies in this way, I need a measure of how incoherent an agent's credence function is. Thus, the fourth section of the paper is devoted to explaining a way in which the degree of incoherence of a credence function can be measured in terms of Dutch books. In the final section of the paper, I use the Dutch book measure of incoherence to evaluate different reasoning strategies non-ideal agents might adopt that mimic rules of reasoning for ideal agent in certain situations. I show that there is an interesting disanalogy between the practical and the epistemic cases: unlike in the practical realm, the best reasoning strategy for an non-ideal agent is always one that mimics an ideal reasoning strategy. However, it is still not very useful advice for a non-ideal agent to pretend to be perfect in choosing a reasoning strategy, because there is usually more than one available "ideal" strategy, and some of strategies that mimic ideal rules can lead to very bad results.

1. Idealization

Ethics, the theory of practical reason, and epistemology have some important commonalities. In particular, they are normative disciplines, which implies that one of their goals is to formulate the norms that govern their respective domains. Ethics and the theory of practical reason are largely concerned with norms governing intentions and actions, whereas epistemology deals with norms governing cognitive states such as belief and levels confidence. Given that all of these disciplines share certain similar goals – formulating norms that govern their respective domains – we can plausibly expect their attempts to do so to exhibit certain parallels with respect to their argumentative strategies, but also with respect to the difficulties they might encounter. In this paper, I want to highlight one of these similar difficulties that come up in both the practical and the epistemological realm, which I will call the idealization problem. It has received considerable attention in ethics and the theory of practical reason but not in epistemology.

Ethical theories and theories of practical reason are in the business of formulating norms that are meant to guide the ways in which we act. Many of these norms are ideal norms, i.e. norms that state what it would take for us to be ideally moral, or ideally practically rational. The problem with ideal norms is that under non-ideal circumstances, it can spell trouble when people try to follow them. We can group the factors that lead to

these difficulties in non-ideal circumstances into two groups: a) the imperfect actions of others, and b) one's own imperfections. Since the imperfect actions of others are not directly relevant to the idealization problem as I discuss it here, I will set them aside for now. (But see Parfit, 2011, Ch. 13 for a discussion of this problem.) I will now explain the latter factor in more detail.

Ordinary agents like us usually are less than perfect in a variety of ways: our capacities for deliberation and decision-making are limited, our attitudes may not be completely coherent, we are biased in a variety of ways, sometimes our emotions get the better of us, and so on. As a result, we encounter difficulties when we try to imitate ideal agents.

One famous example of this kind is the predicament of the angry squash player, which was first introduced by Gary Watson. (Watson, 1975, p. 210) Suppose Ira has just lost a game of squash, and he is very angry about the outcome of the game. What should he do? Ideal Ira would of course be a good loser, and rather than simply walking off, he would go shake his opponent's hand. However, if angry Ira were to walk over to his opponent, he would succumb to his desire to hit her in the face with his squash racquet. Simply walking off seems like a much better option for angry Ira. Thus, we have three possible actions, which we can list from best to worst: walk over and shake hands, simply take off, walk over and hit the opponent in the face. If angry Ira were to try to imitate ideal Ira's best action, he would actually end up producing the worst possible outcome. Thus, rather than trying to be perfect and failing, Ira should protect himself from the bad action of his future self by going with the middle option and simply leaving the squash court.[1] In this case, it is the anticipation of future imperfection that gives rise to the problem. But there are also cases in which an agent's present imperfections lead to similar difficulties.

As Sidgwick has famously discussed, even if we assume that act utilitarianism is the correct moral theory, it is still not obvious whether or not ordinary agents should always apply the utilitarian calculus when deciding how to act. We can find a very concise consideration of this question in Smart (Smart, 1956). He argues that ordinary

[1] Jackson & Pargetter's example of Professor Procrastinate is another instance of this phenomenon. (Jackson & Pargetter, 1986)

agents, because of their limitations and biases, would comply more often with the requirements of morality if they followed common sense moral rules instead of trying to execute utilitarian calculations. He illustrates his point as follows:

> Lack of time is not the only reason why an extreme utilitarian may, on extreme utilitarian principles, trust the rules of common sense morality. He knows that in particular cases where his own interests are involved his calculations are likely to be biased in his own favour. Suppose that he is unhappily married and is deciding whether to get divorced. He will in all probability greatly exaggerate his own unhappiness (and possibly his wife's) and greatly underestimate the harm done to his children by the break up of the family. He will probably also underestimate the likely harm done by the weakening of the general faith in marriage vows. So probably he will come to the correct extreme utilitarian conclusion if he does not in this instance think as an extreme utilitarian but trusts to common sense morality. (Smart, 1956, p. 347)

Smart's example might strike some readers as old-fashioned, but we can easily come up with other examples where one's biases might cause one to misestimate the utility of certain acts. For example, people often overestimate how much pleasure they will get from acquiring certain material possessions, and underestimate the good they could do by donating to charity. Following a common sense moral rule that says that one ought to generally give a certain percentage of one's income to charity might produce the right outcome more reliably in this case.

What examples of this kind show is that, in order for a human agent to approximate a certain ideal norm, being ideally moral in this case, it is not advisable to try to simply imitate an ideally moral agent. An ideally moral agent (from a utilitarian point of view) would be free from biases and self-serving assumptions, so she could simply figure out what is morally required by going through utilitarian calculations. However, for a non-ideal agent to do what is morally best, she should follow a different set of rules, namely those of common sense morality, to get as close as possible to doing what *utilitarianism* requires.

It is not clear, however, whether we can find the same kinds of problems with imitating ideal agents in epistemology. The limitations of real agents are most often brought up in epistemology with the intention of criticizing a particular set of norms for being too demanding. For example, accounts of rational belief that require agents to believe all the logical consequences of their beliefs are often rejected, because agents

simply aren't capable of complying with these norms. Yet, it has not been explored much what happens if a non-ideal agent actually tries to comply with a set of norms that are meant for ideal agents. Based on the examples from the ethics and the theory of practical reason, we should expect there to be a parallel phenomenon in epistemology, namely that agents should not try to imitate ideal agents in their reasoning, because following a different, non-ideal strategy would produce a better outcome. However, I will show that this is not so. In the next section I will begin to explore this issue by briefly characterizing the Bayesian norms on rational degrees of belief.

2. Reasoning with degrees of belief for ideally rational agents

It is commonly assumed that in order to have rational degrees of belief, an agent must have degrees of belief that form a probability function, and she must update her degrees of belief via the conditionalization rule upon learning new evidence.

To set up the probability calculus, we begin with a set of atomic statements $\{A_i\}$, and we will combine it with the standard sentential logical operators to define a language *L*. We will also assume that the relation of logical entailment |= is defined in the classical way. A probability function P on *L* must satisfy the following axioms:

Normalization: For any statement A, if A is a tautology, P (A) = 1
Non-Negativity: For any statement A, P (A) ≥ 0
Finite Additivity: For any two mutually exclusive statements A, B, P (A ∨ B) = P (A) + P (B)
Conditional Probability: For any two statements, P (B|A) = P (B & A) / P (A)

Moreover, there is a rule for updating probabilities in light of new evidence.

Conditionalization: When new evidence A becomes available, the new probability assigned to any statement B is the previous probability of B conditional on A.
So: When A has been added to the body of evidence, $P_{new}(B) = P_{old}(B|A)$ [2]

[2] Some philosophers propose to replace the standard version of the conditionalization principle with Jeffrey Conditionalization. The standard conditionalization rule assumes that when new evidence is learned, it

Many further rules of probability can be derived from the standard axioms, for example Bayes' theorem, and rules for computing the probabilities of logically complex statements. It also follows from the axioms that rational agents must assign the same probability to logically equivalent propositions. The norms on degrees of belief that are derived from the probability axioms are of two kinds: on the one hand, we get norms that govern the relationships between the degrees of belief an agent currently has, and on the other hand, we get norms that govern the adoption of new degrees of belief. We may call the former *coherence norms*, and the latter *update norms*.

The update norms prescribed by the probability axioms dictate what degree of belief to assign to a proposition based on the degrees of belief the agent already has (which may include newly acquired evidence). It is often assumed that the conditionalization rule is the only update norm. However, the other probability axioms can also be viewed as update norms, since they prescribe, for example, how to fill gaps in one's credence function. For example, an agent who has a degree of belief of 0.3 in the proposition that her friend John is currently in San Francisco, and a degree of belief of 0.3 that John is in Los Angeles, may rationally infer a degree of belief of 0.6 in the proposition that John is either in San Francisco or in Los Angeles, as prescribed by the additivity axiom. Similarly for cases in which new evidence is learned: if our agent has a conditional credence of 0.9 in the proposition that John is in Los Angeles, given that John says on facebook that he is in Los Angeles, then it would be rational for him to update his degree of belief in the proposition that John is in Los Angeles to 0.9 upon reading a facebook post by John that says so.

---

becomes certain, i.e. it is assigned probability 1. However, it has often been thought that this condition is too strong, because one doesn't always become certain of evidence propositions. An alternative variant of the conditionalization rule takes this consideration into account, since it allows for evidence that is not certain:

Jeffrey Conditionalization: When an observation bears directly on the probabilities over a partition $\{A_i\}$, changing them from $P_{old}(A_i)$ to $P_{new}(A_i)$, the new probability for any proposition B should be:

$P_{new}(B) = \sum P_{old}(B \mid A_i)\, P_{new}(A_i)$

Thus, Jeffrey conditionalization, unlike standard conditionalization, can accommodate situations where an agent's evidence shifts, but no evidence proposition comes to be known with certainty. Conveniently, standard conditionalization is a special case of Jeffrey conditionalization, where the relevant partition is an evidence proposition and its negation, and the probability of the evidence proposition shifts to 1.

As I mentioned before, the Bayesian norms governing rational degrees of belief are taken to be ideal norms that are not fully realized by humans. Humans fall short of the Bayesian ideal in a variety of ways: first, our degrees of belief don't usually form a coherent probability function. This is in part because we lack the required computational capacities to determine with certainty which propositions are tautologies and contradictions, which means that we will often not be able to assign these propositions the correct extreme credences.[3] We also seem to lack the necessary monitoring capacities to reliably keep our credences coherent. Furthermore, we are prone to making mistakes, so we might introduce errors into our credence function by incorrectly applying the Bayesian rules.

The situation we are in with respect to epistemic norms is thus very similar to the situation we are in with respect to practical norms: the norms characterize an ideal, which is realized by an ideal agent, but not by us. Thus, we may ask: how can we achieve the best results possible for us, given that we aren't ideal agents? Is it a good idea to attempt to imitate ideal agents? We saw that the answer to this question is negative in the practical realm, so we now need to investigate whether the same is true with respect to norms of reasoning.

3. Incoherence and Dutch books

In this section, I will explain my strategy for determining whether it is advisable for non-ideal agents to try to imitate the reasoning strategies used by ideal agents. More specifically, I will assume that a non-ideal agent is an agent whose credences deviate in some way from a coherent probability function. In asking the question whether such an incoherent agent should try to reason according to the same rules as an ideal agent, it is important to recognize that the incoherent agent cannot follow *exactly* the same rules as the ideal agent. If we understand reasoning as the process of forming new credences based on credences that one already has, or updating credences on the basis of new evidence, then the probability axioms don't really say how the agent should go about this, *given* that her credences are incoherent. The probability axioms don't make any straightforward prescriptions about how to reason unless the agent has a coherent

[3] On the problem of assigning extreme credences, see also Christensen, 2007.

credence function as a starting point, which is not true in the cases that we are interested in. However, it is possible to identify certain strategies used by the ideal agent that the incoherent agent can adopt as well, and we can ask what consequences result from the use of these strategies by the incoherent agent.

Before we can do so, we first need to ask: what would constitute a good or bad result of applying a certain reasoning strategy, and how can we determine whether the good or bad result obtains? In the practical cases we discussed in the first section, this was easy to determine. We could easily see that, for example, walking over and hitting your opponent in the face is worse than walking off the squash court, and walking over and shaking your opponent's hand would be better than either of the first two actions. Yet, how can we determine the degree to which a result is good or bad when it comes to a person's credence function? What we need is a way of ranking outcomes of reasoning processes, just like in the practical cases. Here's how this can be done.

Intuitively, agents can be more or less incoherent, depending on how far "off" their credence function is from being a probability function. Consider for example two agents, Sally and Polly, who both only have credences in two propositions, R and ~R. They both have a credence of 0.5 in R. Sally's credence in ~R is 0.49, whereas Polly's credence in ~R is 0.4. Both Sally and Polly have incoherent credences, because their credences in R and ~R don't sum to 1. However, intuitively, Polly's credences are worse than Sally's, because the gap between her actual credences and a coherent credence assignment is much bigger for Polly than for Sally.[4]

Unfortunately, the standard Bayesian framework does not provide us with the resources to distinguish between Polly and Sally, because it merely lets us register the fact that both of them are probabilistically incoherent. The probability axioms by themselves merely define what it takes to have probabilistically coherent credences, but they are silent about any differences between credence functions that diverge from the axioms.

[4] We could give many other examples to motivate this idea, but I don't have room here to discuss them in detail. To see how some other examples would go, consider two agents whose credence functions are the same, except that one of them assigns a credence of 0.99 to some tautology T, whereas the other assigns a credence of 0.5 to T. Intuitively, the latter agent is more incoherent than the former, which is what the measure I develop predicts. We can come up with similar examples involving contradictions.

In our toy example, it is quite intuitive and easy to see which agent is more incoherent. But once we consider agents with larger credence functions, we can't simply see intuitively which agent is more incoherent. What we need is a measure of degrees of incoherence that we can apply to any incoherent credence function, and that will provide us with the rankings that we need. Here, I will focus on a particular kind of measure to do the job – a measure that is based on incoherent agents' vulnerability to Dutch books.

The idea behind Dutch book arguments is to show that an agent whose credence function violates the probability axioms is vulnerable to a guaranteed betting loss from a set of bets that are sanctioned as fair by her credences, whereas a coherent agent faces no such guaranteed loss. The argument rests on the assumption that if one's degree of belief in p is x, then one should consider it fair to buy or sell a ticket for \$xY that pays out \$Y if p is true, and nothing if p is false. Of course, the fact that an agent is prone to losing money is not itself a failure in the epistemic domain. Rather, we should take Dutch books to be dramatizations of certain epistemic failures, which illustrate the problematic commitments that stem from credences that violate certain epistemic norms. (see Christensen, 2004) The incoherent agent is vulnerable to losing money in a Dutch book because she evaluates the same arrangement differently when it is presented in two different ways. Even an agent with incoherent credences would reject as unfair the flat-out suggestion of giving up \$100 for no benefit. Yet an arrangement of bets that leads to the same outcome might be sanctioned as fair by her degrees of belief. By contrast, an agent with coherent credences is not committed to considering a collection of bets as fair that is the same as a request to give away her money for no reason.

On this understanding of Dutch books, it seems very natural to expect differences in possible Dutch book loss to reflect differences in the degree to which credence functions are incoherent. In considering the example of Polly and Sally, we have already gotten a grip on what it means for credence functions to be incoherent to different degrees, so I now want to show how this is reflected in the Dutch books that they license. To avoid unnecessary complications, suppose we are only considering simple agents.[5] They value only money, and value it linearly. Any arrangement in which they give up

[5] I believe that these results also hold, mutatis mutandis, for non-simple agents. However, it would be much more complicated to formulate the arguments for non-simple agents, so sticking to simple agents will do for now.

money for no benefit conflicts with their values. The more money they have to give up for no benefit, the worse they should consider the arrangement to be. As we have already seen, in a Dutch book argument, incoherent credences sanction as fair combinations of bets that guarantee a monetary loss for the agent. It seems very natural to think that the worse the betting arrangement that is sanctioned to be fair, the more incoherent are the agent's credences, and vice versa. That is because the more incoherent the agent's credences are, the greater is the disparity between her two evaluations of equivalents. Thus, if we set up comparable betting arrangements for different incoherent agents, we should expect higher degrees of incoherence to be directly related to greater monetary loss. We can easily see this in the example I introduced above.

Consider Polly and Sally again. Suppose each of them is presented with two tickets. Ticket A says: "Pay $1 to the owner of this ticket if R is true." Ticket B says: "Pay $1 to the owner of this ticket if R is false." Since both of them assign a credence of 0.5 to R, a selling price of $0.50 for ticket A would be fair according to both of them. Given their credences in ~R, a fair selling price for ticket B would be $0.49 for Sally, but $0.40 for Polly. Thus, according to Sally's credences, the sum of the fair selling prices for both tickets is $0.99, whereas according to Polly's credences, it is $0.90. Of course, one of the tickets is guaranteed to win, which means that the seller of the bets must definitely pay out $1 to the buyer. This would leave Sally with a guaranteed loss of $0.01, and Polly with a guaranteed loss of $0.10. Thus, Sally's credences sanction as fair an arrangement that amounts to giving up $0.01 for no benefit, whereas Polly's credences sanction as fair an arrangement that amounts to giving up $0.10 for no benefit. Recall that earlier, we noted that Polly's credences were intuitively more incoherent than Sally's, because they seemed further off from being a probability function. We find this intuitive judgment reflected in the fact that Polly's credences sanction as fair a transaction that leads to a greater monetary loss than Polly's credences in the same betting situation.

We can exploit the fact that the Dutch book loss increases when a person's credences diverge more and more from being probabilistically coherent in order to design a measure of incoherence that utilizes Dutch books. I have designed and defended such a measure in another paper, and I will put it to use here in order to answer the question at

hand. (I prove that the measure has some desirable features in Appendix A.) In the next section, I will briefly explain how the measure works.

4. The maximum Dutch book measure of degrees of incoherence

In this section, I will explain how we can measure degrees of incoherence on the basis of Dutch book loss.[6] In order to use Dutch book loss to measure degrees of incoherence, the measure must fulfill some important conditions. In this section I will propose two basic principles concerning the conditions of adequacy for a Dutch book measure of incoherence. In another paper, I defend these principles in more detail, and I show why the measure I propose is a straightforward implementation of them, but here I want to give at least the basics. The first principle I propose is what I call the *Proportionality Principle*:

> *Proportionality Principle*: A Dutch book measure of a credence function's degree of incoherence should capture our intuitive judgments about degrees of incoherence. Put simply: more incoherence = more Dutch book loss.

The *Proportionality Principle* (PP) is meant to straightforwardly capture the intuitions we have about the relationship between incoherence and Dutch book loss, as exemplified in the case of Sally and Polly.

---

[6] In my paper "Dutch books and Degrees of Incoherence", I discuss different ways of measuring degrees of incoherence, and defend the maximum Dutch book measure. I argue that the maximum Dutch book measure I propose is superior to a different Dutch book measure that has been proposed in a variety of papers by Schervish, Seidenfeld and Kadane. In a nutshell, the problem with their Dutch book measure is that it is ill-equipped to measure an agent's total incoherence. Rather, it measures incoherence by measuring what the "worst" Dutch book is that can be made against the agent. The problem can be illustrated best by aid of a toy example. Suppose there is an agent whose credences are defined over the propositions in the following set: {p, ¬p, q, ¬q}. The agent can adopt one of two credence functions F and G: F(p) = 0.6, F(¬p) = 0.6, F(q) = 0.5, F(¬q) = 0.5, and G(p) = 0.6, G(¬p) = 0.6, G(q) = 0.6, G(¬q) = 0.6. It is easy to see that if the agent adopts F, she is incoherent with respect to her credences in the partition {p, ¬p}, whereas if the adopts G, she is incoherent with respect to her credences in the partition {p, ¬p} and the partition {q,¬q}. Intuitively, if we consider the agent's total incoherence in each case, it seems pretty obvious that adopting G would make the agent more incoherent than adopting F, because if the agent adopts G, there is an additional partition on which she has incoherent credences. However, this is not the result we get from S, S &K's measure. According to their measure, the agent would be equally incoherent in both cases.

A side note is in order here to avoid possible confusion. Of course, the same incoherent credences can give rise to different amounts of Dutch book loss if we allow for variation in the size of the bets involved. For example, a credence of 0.5 in some tautology T sanctions as fair a $1 bet on T that costs $0.50, and also a $2 bet on T that costs $1. The guaranteed loss would be $0.50 for the first bet, and $1 for the second, even though they are based on the same credence. Thus, when I say that more incoherence leads to more Dutch book loss and vice versa, this is meant to be the case in circumstances in which the size of the bets is controlled to ensure the results are commensurable.

Moreover, I take it to be a straightforward consequence of PP that a plausible Dutch book measure should rank credence functions that are probabilistic as being as coherent as each other, and as being more coherent than any non-probabilistic credence function. Lastly, PP ensures that credence functions that can intuitively be compared with each other with respect to their degree of incoherence are part of the same ordering according to any plausible Dutch book measure, i.e. they are not incommensurable with each other.

The second principle I propose as a constraint on acceptable Dutch book measures of incoherence is the *Equality Principle*:

> *Equality Principle*: A Dutch book measure of degrees of incoherence should take all of the agent's degrees of belief into consideration equally, regardless of their content.

In order to see why the *Equality Principle* (EP) is plausible, consider the following example. Suppose K is the proposition that kohlrabis are green, and L is the proposition that lychees are pink. Sally's and Polly's credences are as follows:

Sally:	Polly:
Cr (K) = 0.5	Cr (K) = 0.5
Cr (~K) = 0.4	Cr (~K) = 0.3
Cr (L) = 0.5	Cr (L) = 0.5
Cr (~L) = 0.3	Cr (~L) = 0.4

Intuitively, Sally and Polly have equally incoherent credences. However, if we give more weight to being incoherent with respect to K than to L, then Polly will turn out to have more incoherent credences than Sally, and if we give more weight to L than to K, then Polly will be judged to be more incoherent than Sally. Since neither option seems to be the right result, giving K and L equal weight seems most plausible.

More generally, notice that whether or not a credence function is a probability function has nothing to do with the content of the statements to which credences are assigned. Secondly, remember that we are trying to measure the total incoherence of an agent's credence function. We can only do that if our measure considers every degree of belief the agent has, not leaving anything out. Moreover, as the example illustrates, it seems natural to think that every degree of belief the agent has should be given *equal* consideration in figuring out how incoherent the agent is. As a consequence, the measure should avoid double counting incoherent credences: if an agent is incoherent with respect to some degree of belief of hers, the measure should avoid "punishing" this flaw multiple times.

I will now propose a measure that straightforwardly satisfies both these principles. The basic idea is to set up one giant Dutch book against the agent that covers all of her credences at once, with each bet having fixed stakes of $1. The version of the measure I present here allows for an agent's credence function to have gaps (i.e. it does not need to be defined over a complete Boolean algebra of propositions), but it assumes that agents have the same credence in logically equivalent propositions. Readers who are interested in a more detailed explanation of the measure, and how it can be relaxed to allow for violations of the equivalence rule, should consult my other work on this topic.[7] Here's how the measure works in 5 steps:

1. Take a Boolean algebra BA of n atomic propositions in which the agent S has degrees of belief. More precisely, take a Boolean algebra BA of n atomic propositions, such that for each atomic proposition p in BA, either S has a well-defined credence in p, or p is a constituent of some complex proposition in BA in which S has a well defined credence.

[7] The paper in which I develop the Dutch book measure in detail is called "Dutch books and Degrees of Incoherence". It contains the relevant proofs that show that the measure actually works.

2. Make a list of all the propositions in BA in which the agent has well-defined credences. Let b be the credence function that assigns to every proposition in this list S's credence in this proposition.

3. Then figure out which combination of betting and not betting on or against the propositions on the list will maximize the minimum guaranteed loss across the 2^n state descriptions. For each proposition A_i in which S has a well-defined credence, S can either bet on or against A_i, or not bet on or against A_i.

If S bets on A_i, then S will win $1 - b(A_i)$ if A_i is true and lose $b(A_i)$ if A_i is false.

If S bets against A_i, then S will receive $b(A_i)$ if A_i is false, and lose $1 - b(A_i)$ if A_i is true.

Construct a table in which the rows represent all of the possible combinations of bets for all the propositions in BA, and in which the columns represent the 2^n states of nature.

4. For each combination of bets, identify how much the agent wins or loses given each state description, and determine the amount the agent is guaranteed to lose no matter which state of nature obtains. If there is a state of nature in which the agent gains or breaks even, then the score of this combination of bets is 0, and otherwise the score is the minimum amount lost in any state description.

5. Of all possible combinations of bets, identify the one(s) that yield(s) the highest guaranteed loss across all 2^n states of nature. The guaranteed loss that results from this combination of bets is the degree to which the corresponding credence function b is incoherent.

We can also give the method as one long formula. Suppose b is defined over $\{A_1, \ldots, A_n\}$, and S is the set of all states of nature. Then DOI(b) gives the degree of incoherence for some credence function b.

$$DOI(b) = \max_{\alpha_i \in \{0,1,-1\}} [-\sup_{s \in S} \sum_{i=1}^{n} \alpha_i (I_{A_i}(s) - b(A_i))]$$

This procedure can easily be executed by a computer, even though it obviously becomes very computationally demanding once we reach credence functions of a certain size. However, for our purposes it is sufficient to look at examples of small credence functions, so we will have no problem using the measure. I will now use the maximum Dutch book measure in order to evaluate whether non-ideal agents should use reasoning strategies that are devised for ideal agents.

5. Should non-ideal agents try to use principles of reasoning for ideal agents?

In this section, I will consider the question as to whether non-ideal agents should "pretend they're perfect" and use rules of reasoning that closely resemble the rules that are devised for ideal agents. I will focus on a paradigmatic way of reasoning, in which an agent forms a new attitude based on the attitudes that she already has. More specifically, I consider a particular class of cases in which an agent lacks a credence in some proposition, and she assigns a credence to this proposition based on the other credences she already has. I begin my discussion by looking at a simple example, and I will then draw some more general conclusions.

Suppose there is an agent who has the following credence function:

1. Cr (A&B) = 0.2
2. Cr (A&~B) = 0.2
3. Cr (~A&B) = 0.1
4. Cr (~A&~B) = 0.1
5. Cr (~A) = 0.7
6. Cr (A) = ?

The agent has well-defined credences in the first six propositions in the list, but she does not have a credence assigned to the proposition A. Also, it is easy to see that the credences the agent has are incoherent. Her credences in the first four propositions should sum to 1, but they don't, her credences in 1, 2, and 5 should sum to 1, but they don't, and her in 3 and 4 should sum to the same number as her credence in 5, but they don't. How should the agent go about figuring out what credence to assign to A? Of course, if the agent's other credences were coherent, they would prescribe a unique credence to assign to A, but since the agent is incoherent, this is not the case. However, it seems like the agent can still try to imitate an ideal reasoning strategy by assigning a credence to A that coheres with at least some of her existing credences. If our agent were coherent, each of these strategies would yield the same result, but since she is incoherent, each strategy has a different result. Here are the agent's options for imitating what an ideal agent might do:

a) Since A forms a partition with (~A&B) and (~A&~B), and since the probability axioms prescribe that the sum of one's credences in the propositions in a partition must

be 1, the agent should assign A a credence of 0.8.

b) Since A forms a partition with ~A, and since the probability axioms prescribe that the sum of one's credences in the propositions in a partition must be 1, the agent should assign A a credence of 0.3.

c) Since A is equivalent to (A&B) ∨ (A&~B), and since the probability axioms prescribe that equivalent propositions must be assigned the same credence, the agent should assign A a credence of 0.4.

Thus, there are three "ideal" strategies available to the agent, each of which delivers a different suggestion for assigning a credence to A. There are of course also infinitely many "non-ideal" strategies available to the agent, according to which A is assigned a credence other than 0.8, 0.4 or 0.3. For example, it might be tempting to think that some sort of compromise between the "ideal" credences would deliver the best result. In this case, assigning a credence of 0.55 to A would lie halfway between the two extreme results, so one might think that this could be the optimal credence to assign to A. In order to find out whether the agent would be better off using one of the "ideal", or one of the "non-ideal" strategies for assigning a credence to A, we need to find out how incoherent the agent becomes after employing them. The best strategy for the agent to use is the strategy that minimizes the degree of incoherence of the agent's credence function after assigning a credence to A. If assigning a credence to A according to a "non-ideal" strategy minimizes incoherence, then we have a parallel with the practical case. However, if assigning a credence to A according to one of the "ideal" strategies minimizes incoherence, and this result generalizes, then there is a disanalogy between the theoretical and the practical cases.

To see which strategy is the best, we need to figure out the different ways in which the agent could be Dutch-booked according to our measure, and to see which credence assignment to A results in the lowest Dutch book loss. As I explain in the appendix, in order to set up a maximum Dutch book, we need to be able to divide the propositions in the agent's credence function into subsets of propositions on which the agent ought to be coherent. Those subsets can either be partitions, or anti-partitions[8], or sets that can be

[8] *Definition (Anti-Partition):* Given a partition of propositions PT, the anti-partition PT′ is the set that contains all and only the propositions that are the negations of the propositions in PT. Given that every

further subdivided into two logically equivalent subsets. Then a guaranteed loss will ensure if the agent is not coherent with regard to these subsets. In our example, there are three different ways in which the agent's credences can be divided into subsets, so that a maximum Dutch book can be made against her:

I) {A, (~A&B), (~A&~B)}, {(A&B), (A&~B), ~A}
Loss: |1 – (Cr(A) + 0.2)| + 0.1
II) {A, (A&B), (A&~B)}, {(~A&B), (~A&~B), ~A}
Loss: |Cr(A) – 0.4| + 0.5
III) {A, ~A}, {(A&B), (A&~B), (~A&B), (~A&~B)}
Loss: |1 – (Cr(A) + 0.7)|+ 0.4

It is easy to see here that the agent is already guaranteed to lose in a maximum Dutch book even before she assigns a credence to A. The maximum guaranteed loss, i.e. the degree to which she is incoherent according to our measure, before assigning any credence to A is 0.5, as we can see from option II). Since adding a new credence to an existing credence function can never decrease an agent's degree of incoherence, the best the agent can do in assigning a credence to A is to maintain a degree of incoherence of 0.5. Adding a new credence can never decrease one's degree of incoherence since in that case, one of the sub-Dutch books that are part of the maximum Dutch book would have to produce a negative degree of incoherence, which is impossible.

Is there any credence assignment to A that will lead to no increase in the agent's degree of incoherence, leaving it at 0.5? It turns out that there is. If the agent assigns Cr(A) = 0.4, then no matter which one of the options I), II) and III) we choose for the Dutch book, her degree of incoherence remains 0.5. If we choose any other credence assignment for A, then her degree of incoherence will be above 0.5, leading to a worse result. This means that the best reasoning strategy for the agent in this example is one of the ideal strategies, namely strategy c), where the agent assigns to A the same credence as the sum of her credences in (A&B) and (A&~B). Yet, we can also see that the other two

---

partition contains exactly one true proposition, every anti-partition correspondingly contains exactly one false proposition.

proposed "ideal" strategies a) and b) lead to rather bad results in comparison. If the agent assigned A a credence of 0.8 according to strategy a), or a credence of 0.3 according to strategy b), then her resulting degree of incoherence would be 0.9 or 0.6 respectively, which is even worse than the degree of incoherence that some of the non-ideal strategies would deliver. A non-ideal strategy that assigns A a credence closer to 0.4 would deliver a better result. The non-ideal "compromise" strategy, which would assign A a credence of 0.55 would result in a degree of incoherence of 0.65, which is better than strategy a), but worse than strategy b). Thus, we find that in this example, there is an "ideal" strategy that the agent can choose that will lead to the best outcome of her reasoning, but she cannot choose just any of the available perfect strategies to achieve this result.

It turns out that the result we have seen in this example generalizes. Take any set of propositions that can be divided into two subsets in such a way that each subset is either a partition, an anti-partition, or can be further divided into two equivalent subsets. Suppose that there is an agent who has credences assigned to all but one of the propositions in this set, and who is trying to determine the best way to assign a credence to the remaining proposition, call it A. It turns out that there is always a way of assigning a credence to A that does not increase the degree to which the agent was incoherent before assigning a credence to A. In order to determine what credence to assign to a, the agent must determine which division of the propositions of her credence function into subsets leads to the highest guaranteed loss prior to assigning a credence to A. (In our example, this was option II.) The credence the agent must assign to A in order to not increase her degree of incoherence is the one that is suggested by this division of her credences into subsets (for a proof of this, see appendix B). In other words, once the agent has determined which way of dividing up the propositions in her credence function into subsets leads to the highest guaranteed loss prior to assigning a credence to A, she knows that A needs to be assigned the credence that makes it coherent with the other credences in the subset that A is part of. This way, adding a credence in A to her credence function will not increase the agent's previous degree of incoherence.

And since making one's new credence coherent with (at least part of) one's existing credences is how we conceived of a reasoning strategy that imitates what an ideal agent would do, we can conclude that the best reasoning strategy for the agent is indeed one

that imitates an ideal agent.

The result as is presented here is proved only for relatively small credence functions, but evidence from looking at larger credence functions, where there are more than 2 appropriate subsets into which we can divide the propositions in the agent's credence function, suggests that this result holds more generally.

Does this mean that incoherent agents should pretend they're perfect when trying to assign a credence to a proposition based on their existing credences? We have shown that one of the "ideal" strategies is always going to be the optimal strategy for the agent, because it adds no additional incoherence to her credence function. However, the problem is that the agent can't just pick any one of the "ideal" strategies, because some of them can result in very bad increases in incoherence. In order to know which ideal strategy is the correct one, which way of setting up a Dutch book against her prior to assigning the new credence will maximize her degree of incoherence, because this is the information she needs in order to know which subset of her credence function she must make her new credence coherent with. This knowledge is not easy to come by. It is an open question whether there might be some heuristic the agent could apply that would relatively reliably predict which one of the ideal strategies is the correct one. If there were such a heuristic, then there would be an easy way for the agent to find out how to assign a credence to A. However, if there is no such heuristic, it might be too complicated for the agent to figure out how to pretend to be perfect for this to be a very useful reasoning strategy.

Conclusion

I began this essay by asking the question of whether non-ideal agents should try to emulate ideal agents in order to comply better with ideal norms. I pointed out that in the literature on practical norms, several authors have argued that it is a bad strategy for non-ideal agents to imitate ideal agents, because the results of this can be worse than if the non-ideal agents didn't try to imitate an ideal agent. I then wondered whether the same was true with respect to ideal norms and agents in the epistemic realm. I explained how a Dutch book measure can be used to determine an agent's degree of incoherence, and I argued that by aid of this measure we can find out whether it is acceptable for a non-ideal

agent to imitate the reasoning strategies of an ideal agent. Focusing on relatively small credence functions, I showed that agents usually have multiple different ways in which they can try to imitate an ideal agent in assigning a new credence to a proposition. It turns out that one of these "ideal" strategies is the optimal reasoning strategy for the agent to follow. This is an interesting disanalogy with the practical cases, because we found that there are practical cases in which no ideal strategy is better than some non-ideal strategy. However, we also saw that some "ideal" reasoning strategies lead to very bad results, i.e. significant increases in incoherence, and that it is very difficult for the agent to know which ideal reasoning strategy to imitate. In order for the agent to know which ideal strategy is the right one, she must know how the optimal Dutch book against her is set up prior to her assigning the new credence.

We can draw two morals from this: first, the practical and the epistemic cases are different: the best strategy is the epistemic case is always one that imitates ideal reasoning in some way, whereas this is not true in the practical case. Second, we can still not advise non-ideal agents to simply "pretend to be perfect", because some of the ways of imitating "ideal" reasoning strategies lead to bad results. Yet, finding the correct way of pretending to be perfect is complicated. So, unless there is a heuristic for identifying the correct way of pretending to be perfect, this strategy might be too complicated to be helpful for non-ideal agents.

Bibliography

Christensen, David (2007): "Does Murphy's Law Apply in Epistemology?", in: *Oxford Studies in Epistemology*, Vol. 2, pp. 3-31.

Christensen, David (2004): *Putting Logic in its Place*, OUP.

Hájek, Alan (2003): "What Conditional Probability Could Not Be", in: *Synthese*, Vol. 137, pp. 273-323.

Jackson, Frank & Pargetter, Robert (1986): "Oughts, Options, and Actualism", in: *The Philosophical Review*, Vol. 95, No. 2, pp. 233-255.

Kant, Immanuel (1999): *Grundlegung zur Metaphysik der Sitten*, 7th ed., Meiner.

Parfit, Derek (2011): *On What Matters*, Vol. 1, Oxford University Press.

Schervish, Mark J; Seidenfeld, Teddy & Kadane, Joseph B. (2000): "How Sets of Coherent Probabilities May Serve as Models for Degrees of Incoherence", in: *International Journal of Uncertainty, Fuzziness and Knowlegde-based Systems* 8, pp. 347-355.

Schervish, Mark J; Seidenfeld, Teddy & Kadane, Joseph B. (2002a): "Measuring Incoherence", in: *Sankhya: The Indian Journal of Statistics*, Vol. 64, Series A, Pt. 3, pp. 561-587.

Schervish, Mark J; Seidenfeld, Teddy & Kadane, Joseph B. (2002b): "Measures of Incoherence: How not to Gamble if You Must", in: Bernardo, J. M., et. al. (ed.): *Bayesian Statistics 7*, Oxford University Press, 2003, pp. 385-401.

Smart, J.J.C. (1956): "Extreme and Restricted Utilitarianism", in: *The Philosophical Quarterly*, Vol. 6, No. 25, pp. 344-354.

Watson, Gary (1975): "Free Agency", in: *Journal of Philosophy*, Vol. 72, No. 8, pp. 205-220.

Weisberg, Jonathan (2011): Varieties of Bayesianism. in: Gabbay, D.; Hartmann, S. & Woods, J. (eds.): *Handbook of the History of Logic, Vol. 10*. Elsevier.

Appendix A:

How the maximum Dutch book measure works:

To see how the maximum Dutch book measure works, it is easiest to first consider cases in which the agent's credence function is a complete Boolean algebra of propositions. I will first prove that the measure works for this case, and then introduce the possibility of having gaps in one's credence function in a second step. Here's the method for the non-gappy case:

1. Take a Boolean algebra BA of n atomic propositions in which the agent S has degrees of belief.

2. Make a list of all the propositions in BA and let b be the credence function that assigns to every proposition in this list S's credence in this proposition.

3. Then figure out which combination of betting on or against the propositions on the list will maximize the minimum guaranteed loss across the 2^n state descriptions. For each proposition A_i in BA, S can either bet on or against A_i.

If S bets on A_i, then S will win $1 - b(A_i)$ if A_i is true and lose $b(A_i)$ if A_i is false.
If S bets against A_i, then S will receive $b(A_i)$ if A_i is false, and lose $1 - b(A_i)$ if A_i is true.
Construct a table in which the rows represent all of the possible combinations of bets for all the propositions in BA, and in which the columns represent the 2^n states of nature.
4. For each combination of bets, identify how much the agent wins or loses given each state description, and determine the amount the agent is guaranteed to lose no matter which state of nature obtains. If there is a state of nature in which the agent gains or breaks even, then the score of this combination of bets is 0, and otherwise the score is the minimum amount lost in any state description.
5. Of all possible combinations of bets, identify the one(s) that yield(s) the highest guaranteed loss across all 2^n states of nature. The guaranteed loss that results from this combination of bets is the degree to which the corresponding credence function b is incoherent.

In order to show that this method works, I need to show that an agent is maximum-Dutch-bookable if and only if the agent is incoherent. Clearly, if an agent is maximum Dutch-bookable, then the agent is Dutch-bookable, and thus incoherent. I need to show that if an agent is incoherent, then she is maximum Dutch-bookable. This can be shown on the basis of three lemmas:

Lemma 1: If an agent is incoherent then there is some partition on which she is incoherent.[9]
Proof: Suppose the agent is probabilistically coherent on all partitions of her credence function b. Now suppose for reductio that the agent is non-probabilistic on some subset S of her credence function b that is not a partition. The only ways of being non-probabilistic on S are by violating one or more of the three axioms. Consider each axiom in turn.

[9] To be incoherent on a partition of propositions means that there is a partition of propositions to which the agent assigns credences, and these credences violate the probability axioms. If we only allow for credences between 0 and 1, then being incoherent on a partition means that the credences that the agent assigns to the propositions in the partition don't sum to 1. If we want to allow for negative credences, and credences greater than 1, then an agent can also be incoherent by assigning a credence that is not between 0 and 1 to any proposition, regardless of whether the credences in the partition sum to 1.

Normality: Assume the agent assigns a probability other than 1 to the tautology T in her credence function. By assumption, the agent is probabilistically coherent on all partitions of her credence function, including the partition that contains T. This is a contradiction.
Non-Negativity: Assume the agent assigns a negative probability to some proposition A in her credence function. By assumption, the agent is probabilistically coherent on all partitions of her credence function, including the partition {A, ~A}. This is a contradiction.
Finite additivity: Assume there exists a pair of propositions {A, B} that are mutually exclusive, but to which the agent assigns credences such that b(A v B) ≠ b(A) + b(B). Consider the following two partitions: PT_1 = {A, B,~A&~B}, and PT_2 = {A v B, ~A&~B}. By assumption, b must be probabilistic on both PT_1 and PT_2. This means that the following two constraints must be satisfied by b: (i) b(A) + b(B) + b(~A&~B) = 1, and (ii) b(A v B) + b(~A&~B) = 1. This contradicts b(A v B) ≠ b(A) + b(B).

Lemma 2: If there exists a partition with respect to which an agent is incoherent, then the Boolean algebra is divisible into partitions and anti-partitions, such that she is incoherent with respect to at least one of them.
Proof: For this proof, we first need the definition of an anti-partition:

> *Definition (Anti-Partition):* Given a partition of propositions PT, the anti-partition PT′ is the set that contains all and only the propositions that are the negations of the propositions in PT. Given that every partition contains exactly one true proposition, every anti-partition correspondingly contains exactly one false proposition.

Remember that a Boolean algebra of propositions is structured in such a way that for each proposition it contains, it also contains the negation of that proposition. Assume that there is some partition on which the agent is incoherent. We can now identify the anti-partition that corresponds to the partition on which the agent is incoherent. Since this anti-partition contains exactly the propositions that are the negations of the propositions in the partition on which the agent is incoherent, the part of the Boolean algebra that is

encompassed by the resulting partition/anti-partition duo does not contain any proposition without also containing its negation. The now remaining part of the credence function can be divided into partitions by simply pairing each remaining proposition with its negation.

Lemma 3: If the Boolean algebra is divisible into partitions and anti-partitions such that the agent is incoherent with respect to at least one of them, then she is maximum Dutch-bookable.

Proof: In order to prove lemma 3, we need to show two things: First, we need to show that any partition or anti-partition on which the agent is incoherent will generate a Dutch book loss, given that all the bets have fixed $1 stakes. Secondly, we need to show that for all of the remaining partitions or anti-partitions on which the agent is coherent, the bookie can set up the $1 bets in such a way that the agent neither gains nor loses money. This is important, because it guarantees that there won't be any accidental gains from these bets that decrease the guaranteed loss from the bets on the incoherent parts of the credence function.

1) Being incoherent on any partition or anti-partition leads to a guaranteed Dutch book loss, given that all bets have $1 stakes.

Partition:

Case A (the credences don't sum to 1): Suppose the agent has credences in the propositions in a partition that don't sum to 1. If her credences sum to some number $x<1$, her credences justify a combined selling price of $x for the bets on the propositions in the partition. Since exactly one proposition in the partition is true, she will be forced to pay out $1 for the winning bet, leaving her with a net loss of $x-1. If her credences sum to some number $y>1$, her credences justify a combined buying price of $y for the bets on the propositions in the partition. Since exactly one proposition in the partition is true, she will gain $1 for the winning bet, leaving her with a net loss of $1-y.

Case B (the credences sum to 1, but are not individually coherent):

Suppose an agent has credences in the propositions in a partition that sum to 1, yet the credences are not individually coherent, since at least one credence is negative. Let x represent the sum of the credences that are negative, and y the sum of the credences that

are positive. Thus, $x<0$ and $y>0$ and $x+y=1$. We can Dutch book the agent who has these credences by making her sell the bets on propositions in which she has negative credences and buy the bets on propositions in which she has positive credences. Selling a bet in which one has a negative credence means that one "sells" the bet for nothing and gives the buyer some money on top of it that is equal to the absolute value of one's credence. So if, the sum of the agent's negative credences is x, she will "sell" the corresponding bets for $x, meaning she has spent $–x. She will also buy the bets on her positive credences for $y, meaning that she spends a total of $-x+y. And since $-x+y>1$, she spends a total of more than $1. Depending on which one of the bets wins, she either gains back $1 or has to pay out $1. That means that she can't win back enough money to cover her expenses, leaving her with a guaranteed loss.
The proof works analogously for credences that are greater than 1.

Anti-Partition:

Case A (the credences don't sum to n-1): Suppose the agent has credences in the propositions in an anti-partition with n members that don't sum to (n-1). If her credences sum to some number $x<(n-1)$, her credences justify a combined selling price of $x for the bets on the propositions in the partition. Since exactly one proposition in the partition is false, she will be forced to pay out $n-1 for the winning bets, leaving her with a net loss of $x-(n-1). If her credences sum to some number $y>(n-1)$, her credences justify a combined buying price of $y for the bets on the propositions in the partition. Since exactly one proposition in the partition is false, she will gain $(n-1) for the winning bets, leaving her with a net loss of $(1-n)-y.

Case B (the credences sum to (n-1), but are not individually coherent):
Suppose an agent has credences in the propositions in an anti-partition with n members that sum to (n-1), yet the credences are not individually coherent, since at least one credence is negative. Let x represent the sum of the credences that are negative, and y the sum of the credences that are positive. Thus, $x<0$ and $y>0$ and $x+y=n-1$. We can Dutch book the agent who has these credences by making her sell the bets on propositions in which she has negative credences and buy the bets on propositions in which she has

positive credences. Selling a bet in which one has a negative credence means that one "sells" the bet for nothing and gives the buyer some money on top of it that is equal to the absolute value of one's credence. So if, the sum of the agent's negative credences is x, she will "sell" the corresponding bets for $x, meaning she has spent $–x. She will also buy the bets on her positive credences for $y, meaning that she spends a total of $-x+y. And since -x+y>(n-1), she spends a total of more than $(n-1). The maximum amount of money she can win back is $(n-1), leaving her with a guaranteed loss.
The proof works analogously for credences that are greater than 1.

2) If the agent is coherent on a partition or anti-partition, there is a combination of $1 bets that leads to no gain or loss for the agent.

Partition:

If the agent has coherent credences in a partition of propositions, then her credences in all the members of the partition are individually coherent, and they sum up to 1. That means that her credences justify a combined selling or buying price of $1 for all the bets on the propositions in the partition, given that all the bets have $1 stakes. Since exactly one proposition in the partition must be true, exactly one of the bets in the package will win. Thus, if the agent buys all the bets for $1, she will win $1 back, with a net gain/loss of $0. Similarly, if the agent sells all bets for a combined price of $1, she will have to pay out $1 for the winning bet, leaving her with a net gain/loss of $0.

Anti-partition:

If the agent has coherent credences in an anti-partition of propositions that has n members, then her credences in all the members of the anti-partition are individually coherent, and they sum up to (n-1). That means that her credences justify a combined selling or buying price of $(n-1) for all the bets on the propositions in the anti-partition, given that all the bets have $1 stakes. Since exactly one proposition in the anti-partition must be false, exactly n-1 of the bets will be winning bets. Thus, if the agent buys all the bets for $n-1, she will win $n-1 back, with a net gain/loss of $0. Similarly, if the agent sells all bets for a combined price of $n-1, she will have to pay out $n-1 for the winning bets, leaving her with a net gain/loss of $0.

I have shown that the partitions or anti-partitions on which the agent is incoherent will produce a guaranteed loss, whereas the partitions or anti-partitions on which the agent is coherent produce zero gain or loss, and therefore don't mask the guaranteed loss resulting from the agent's incoherent credences. I have therefore shown that if the Boolean algebra is divisible into partitions and anti-partitions such that the agent is incoherent with respect to at least one of them, then she is maximum Dutch-bookable. Thus, I have proved all three lemmas, from which it follows that any agent with incoherent credences is maximum Dutch-bookable.

Fixing the bet size:

As I explain in my discussion of Dutch book measures, any Dutch book measure of degrees of incoherence must be normalized, so as to avoid that an agent counts as more incoherent than another simply because the bets in her Dutch book are larger. In footnote 6 I argue that the ways in which Schervish, Seidenfeld & Kadane set up and normalize their incoherence measure does not capture the agent's total incoherence, which is why I propose a different Dutch book measure that has the normalization directly built into the size of the bets. I propose to normalize the Dutch books by setting the stakes of each bet to $1. In other words, for each bet, the amount that can be gained and the amount that can be lost in the bet sum to $1. For example, if Sally has a credence of 0.8 in p, the corresponding bet on p would cost Sally $0.80, and she would be paid $1 if p were true, resulting in a net gain of $0.20 for her. The choice of making each bet have a stake of $1 is purely arbitrary; it would not make any difference if we chose any other positive amount.

However, there are many other ways in which one might be able to normalize the bets. Here, I discuss three other options for fixing the size of a bet that might seem natural, and I show why my method is preferable to these alternatives. The first possibility is to make all bets cost $1, so that the agent's buying and selling price for each bet is constant, and the payoff varies with the agent's credence. The second possibility is to make the payoff of each bet be $1, so that the net amount the agent can gain if she wins the bet is constant. The third possibility is to make the product of the possible net gain and net loss equal to $1 for every bet.

The problem that arises for these three possibilities, but not for my chosen normalization, is that they cannot handle bets on or against propositions to which the agent assigns credences of 1 or 0.

Suppose Sally assigns a credence of 0 to some proposition q. According to my chosen normalization, this means that Sally would only accept a bet on q if she could bet for free, winning $1 if p turns out to be true. If we choose the first alternative, making each bet cost $1, we cannot capture the bet Sally would be willing to make in this case. The payoff, or net gain for a bet that costs $1 is (1-1/P(q), which is undefined if Sally's credence is 0. Or, in other words, if Sally is only willing to accept bets on q that are free, this is incompatible with stipulating that all the bets she is willing to accept cost $1. Thus, there are certain acceptable bets that are incompatible with this normalization.

If we choose the second alternative, making the net gain from each bet $1, we get a similar problem. Suppose Sally assigns credence 1 to some proposition r. According to my chosen normalization, this means that Sally is willing to pay $1 for a bet that pays nothing if r is correct. But according to the second alternative, there cannot be any bets that involve a net gain of $0, because it stipulates that every bet that Sally is willing to accept has a net gain of $1. Thus, there are bets that Sally is willing to make that cannot be captured by this normalization. Alternatively, if we made the bet cost $1, it might mean that Sally would be willing to pay any price for it, which would also be problematic for the measure, since in that case her credence would not determine a price for the bet.

A further problem with the two options just discussed is that they do not guarantee that there is a Dutch book to be made in every case in which an agent has incoherent credences. For example, if an agent has credence 0.8 in some proposition p, and 0.1 in ¬p, neither one of the two previous normalizations leads to a guaranteed loss.

If we choose the third alternative, we stipulate that the product of the net loss and the net gain of any bet must equal $1. This is obviously incompatible both with bets that cost $0 and with bets that have a net gain of $0, because in each case the product of net loss and net gain would be 0. However, since both of these types of bets can occur, the third alternative is not a suitable normalization either.

Thus, we are left with my chosen normalization, which assumes that the sum of the net loss and net gain for each bet is $1. This normalization does not have a problem with bets based on credence assignments of 0 or 1.

Gappy credence functions

It is not difficult to see that the maximum Dutch book measure can handle gappy credence functions. An agent has a gappy credence function just in case her credences are not defined over a full Boolean algebra. That means that there are propositions that are part of the Boolean algebra, but that the agent has not assigned a credence to, for example because she has never even considered that particular proposition. Given that the maximum Dutch book measure relies on the fact that a Boolean algebra can be divided up into partitions and anti-partitions, the gaps will affect the measure insofar as there are now certain partitions and anti-partitions such that the agent only has credences in some of the propositions in the relevant sets. For example, suppose the partition PT = {p, ¬p&q, ¬p&¬q} is part of the Boolean algebra that the agent's credence function *would be* defined over if it *weren't* gappy. If the agent's credence function is gappy with respect to PT, this means that the agent has not assigned a credence to one or more of the propositions in PT. Similarly for anti-partitions. The main problem with gappy credence functions is that we cannot rely on the agent's coherent credences to cancel each other out in a maximum Dutch book. There might be individual propositions in which the agent has a credence that are such that if a bet on or against that proposition is included in the maximum Dutch book, the bet could generate a gain in some state descriptions that masks the guaranteed loss the agent faces because of her incoherent credences.

We can modify the measure to circumvent this problem by requiring that instead of there being two options for each proposition in which the agent has a credence – betting on it or betting against it – there are now three options: betting on the proposition, against it, or leaving it out of the Dutch book. This modification ensures that all the propositions that lead to Dutch book loss will be included in the maximum Dutch book, but that propositions that could generate accidental gains can be left out.

Case 1: The agent has assigned credences to none of the propositions in some partition or anti-partition

In this scenario, we don't encounter any problems with the measure, since the entire partition or anti-partition can simply be left out of the maximum Dutch book.

Case 2: The agent has assigned credences to some but not all propositions in a partition or anti-partition, and the assigned credences are incoherent

A) Partitions:

There are two possibilities here: Either some of the propositions in the partitions are assigned credences that sum to some number x greater than 1 or below 0, or at least one of the credences is not between 0 and 1. Let's consider these cases in turn.

In the first case, we can make the agent buy (if $x>1$) or sell (if $x<0$) the bets on the relevant propositions for a combined price of $x. If she buys the bets, she can win maximally $1 back, leaving her with a guaranteed loss. If she sells the bets for some amount $x<0$, she will lose at least $-x.

In the second case, suppose the agent has a credence in one of the propositions in the partition that is not between 0 and 1, and she also assigns a coherent credence to one or more other propositions in the partition, and the sum of all these credences is between 0 and 1. In this case, we can definitely make the agent bet on or against the propositions in which she has credences that are individually incoherent (not between 0 and 1). If her credence in some proposition is greater than 1, we can make her buy the bet for the price justified by her credence, and since she will be able to win back no more than $1, she will face a guaranteed loss. If her credence in some proposition is below 1, we can make her sell the bet for the price justified by her credence, which means that she will lose at least the amount she has to give to the seller.

However, it may be necessary to exclude the proposition in which she has a coherent credence from the maximum Dutch book in this case, because the possible gains from these bets can mask the guaranteed losses justified by the individually incoherent credences in some cases. For example, suppose two propositions p and q are mutually exclusive, but not jointly exhaustive. The agent lacks credences in any other propositions. And suppose the agent assigns a credence of 0.8 to p and a credence of -0.1 to q. In this case, no Dutch book can be made that involves bets on both p and q, we can only make a

Dutch book if the agent bets only on q. However, the maximum Dutch book measure automatically considers the difference between including and not including each proposition in the maximum Dutch book whenever a credence function is gappy. It is therefore guaranteed that a combination of bets that includes bets that would lead to gains that mask guaranteed Dutch book losses from other propositions will be ruled out as the combination of bets that determines the agent's degree of incoherence.

B) Anti-Partitions:

There are two possibilities here: Either some of the propositions in the anti-partitions are assigned credences that sum to some number greater than (n-1) or below 0, or at least one of the credences is not between 0 and 1. Let's consider these cases in turn. In the first case, we can make the agent buy (if $x>(n-1)$) or sell (if $x<0$) the bets on the relevant propositions for a combined price of $x. If she buys the bets, she can win maximally $(n-1) back, leaving her with a guaranteed loss. If she sells the bets for some amount $x<0$, she will lose at least $-x.

In the second case, suppose the agent has a credence in one of the propositions in the anti-partition that is greater than 1 or below 0, and she also assigns a coherent credence to one or more other propositions in the anti-partition, and the sum of all these credences is between 0 and (n-1). Again, it may be necessary to leave the propositions in which the agent has coherent credences out of the maximum Dutch book, for the reasons explained before. As before, the setup of the measure takes care of this.

Case 3: The agent has assigned credences to some but not all propositions in PT, and the assigned credences are by themselves coherent

In this case, like in some of the cases discussed in the previous section, it may be necessary to exclude bets on the relevant propositions from the maximum Dutch book. If an agent has *non-gappy*, coherent credences in the propositions in a partition or anti-partition, the maximum Dutch book method ensures that the bets on or against these propositions cancel each other out, thus leading to neither a gain nor a loss. However, if an agent has *gappy*, coherent credences in the propositions in a partition or anti-partition, then including these bets in a maximum Dutch book can in certain distort the measure of

the agent's degree of incoherence. That is because some worlds are such that the agent would win money from making these bets, which would diminish the guaranteed loss the agent faces in virtue of her incoherent credences. For this reason, bets on propositions that are part of gappy partitions or anti-partitions must sometimes be excluded from the maximum Dutch book if those credences are coherent. Whether or not these propositions must be excluded can be seen very easily by looking at the different possible combinations of bets, and by determining which combination leads to the greatest guaranteed loss across all state descriptions.

In sum, we can easily change the original measure to account for the effects of gappy credence functions by allowing that certain propositions may be left out of the Dutch book. Thus, we can reformulate Steps 1-5 of setting up the measure as I do in the main body of the paper.

Notice that the resulting measure still complies with the Proportionality Principle and the Equality Principle. The only significant change we've made to the measure to account for gappiness is to allow that certain coherent credences can be left out of the Dutch book, so that potential gains from bets on these credences don't mask the guaranteed loss from the agent's incoherent credences. The agent's incoherent credences are still taken into account by the measure in the exact same way as before.

Appendix B

Insert proof here that shows how it follows from the triangle inequality that assigning a credence according to the procedure described in the main text never leads to an increase in incoherence.