Salience and the Sure-Thing Principle

Chloé de Canson

Abstract. The Two-Envelope Paradox is a counterexample to the Sure-Thing Principle. In this paper, I isolate the feature of the Sure-Thing Principle which generates this paradox, and I formulate a restricted version of the principle which defuses the paradox. The analysis leads to a surprising result; namely, that the Sure-Thing Principle is relative to the way in which possibilities are described, in a way I make precise. This has important implications for decision theory: it implies that there is an additional layer of description-relativism in the individuation of decision-theoretic situations, beyond standard intentionality.

1.

This paper is about what is known as the ***Sure-Thing Principle***, which originates with Savage (1954). Savage introduces this principle with the help of the following case: 'A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he knew that the Republican candidate were going to win, and again finds that he would. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains' (p. 21). The intuition behind this case can be captured by the following principle. Where $\Omega = \{w_1, w_2, ...\}$ is the set of all the states an agent considers possible, or *sample space*:

> **Unrestricted Sure-Thing Principle (STP$_\text{U}$).** Let $\mathcal{P} = \{P_1, P_2, ...\}$ form a partition of Ω. If, for every proposition $P_i \in \mathcal{P}$, an agent would perform an action T if she were to learn P_i, then the agent ought to perform T.

In this paper, I examine a counterexample to the (STP$_\text{U}$): the *paradox of the two envelopes.* My aim is to isolate the feature of the (STP$_\text{U}$) which generates this paradox, and to formulate a restricted version of the Sure-Thing Principle, which both captures the intuition behind Savage's businessman case, and avoids the paradoxical conclusion. The analysis leads to a surprising result; namely, that the Sure-Thing Principle is relative to the way in which possibilities are described, in a way I make precise. This has important implications for decision theory: it implies that there is an additional layer of description-relativism in the individuation of decision-theoretic situations, beyond standard intensionality.

I proceed in the following way. I begin, in §2, by presenting the paradox of the two envelopes, and showing that it is a counterexample to the (STP_U). Then, in §3, I argue that a refinement of the (STP_U) proposed by Dietrich and List (2005) is unsatisfactory. In §4, I use my criticism of Dietrich and List to draw a structural analogy between the two-envelope paradox and the grue paradox, and I introduce the importance of a specific kind of description-relativism for the applicability of the Sure-Thing Principle, called *salience*. This leads to a refinement of the (STP_U), in §5, which I show solves the paradox.

2.

There are two envelopes, A and B, in front of you. You are told that they each contain some amount of money from the set $\{1, 2, 4, 8, ...\}$, and that one contains an amount twice as large as the other, but you do not know which. You do know that, for any given amount of money, it is equally likely to be in A and in B. You also know, for any given amount of money, how likely it is for that amount to be in one of the two envelopes.

You represent the situation as follows. The sample space $\Omega = \{w_1, w_2, ...\}$ is the set of all possible states, that is, all possible contents of both envelopes. Random variables A and B are defined on Ω, representing for each state the amount you get in that state if you get envelope A or envelope B, respectively. So for instance, if w_1 is the state where A contains 1 and B contains 2, $A(w_1) = 1$ and $B(w_1) = 2$. Given that the sample space is not finite, because the set of possible amounts in each envelope is not finite, the probability distribution over Ω cannot be uniform. So, in this paper, I will be assuming that the probability distribution is the one proposed by Broome (1995), and I call this distribution p.[1] There may be other probability distributions from which a two-envelope paradox can be generated, but this does not matter for my dialectical purposes: a single counterexample to (STP_U) is sufficient. All this can be represented in the following matrix **1**.

[1] All technical details can be found in Broome's paper. In a nutshell, his proposal is that, for any $k \in \{0, 1, 2, 3, ...\}$, the probability of finding some amount in an envelope is determined by the joint distribution $p(A = 2^k \wedge B = 2^{k+1}) = p(B = 2^k \wedge A = 2^{k+1}) = \frac{2^{k-1}}{3^{k+1}}$.

	w_1	w_2	w_3	w_4	w_5	$\cdots$
p	$1/6$	$1/6$	$1/9$	$1/9$	$2/27$	$\cdots$
A	1	2	2	4	4	$\cdots$
B	2	1	4	2	8	$\cdots$

Matrix 1

Now, onto the decision problem. You are given envelope A, and you are offered the opportunity to switch to envelope B. Let us call this the *initial case*. What should you do?

In order to determine whether you should switch to B, you reason in the following way. If envelope A contains 1, that is, if event $\{w_1\}$ obtains, then envelope B contains 2, and you should switch. If envelope A contains 2, that is, if event $\{w_2, w_3\}$ obtains, then envelope B contains 1 with probability $3/5$, and 4 with probability $2/5$.[2] So the expected utility of switching to envelope B, given that envelope A contains 2 is $E[B|A = 2] = 1 \cdot 3/5 + 4 \cdot 2/5 = 11/5 > 2$. So, if envelope A contains 2, you should also switch. In fact, it can be shown that the amount in envelope B conditional on the amount in A is:[3]

$$E[B|A = x] = \begin{cases} 2x & \text{if } x = 1 \\ 11x/10 & \text{if } x > 1 \end{cases}$$

So $E[B|A = x] > E[A|A = x]$, for any x. In other words, no matter what the amount in your envelope is, the expected amount in evelope B is higher, and you ought to switch to envelope B. Therefore, you reason, you ought to switch to envelope B.

This is counter-intuitive, but not yet paradoxical. However, a paradox arises when we consider the following analogous reasoning. If envelope B contains 1, that is, if event $\{w_2\}$ obtains, then envelope A contains 2, and you should switch. If envelope B contains 2, that is, if event $\{w_1, w_4\}$ obtains, then envelope A contains 1 with probability $3/5$, and 4 with probability $2/5$. This means that $E[A|B = 2] = 1 \cdot 3/5 + 4 \cdot 2/5 = 11/5 > 2$. So, if envelope B contains 2, you should also switch. In fact, it can be shown that conditional on the amount in B, the amount in A is:[4]

$$E[A|B = y] = \begin{cases} 2y & \text{if } y = 1 \\ 11y/10 & \text{if } y > 1 \end{cases}$$

[2] See Broome (1995) for details on why the ratios are $2/5$ and $3/5$ on distribution p—though you can see by looking at matrix **1** that p is such that $p(w_2) > p(w_3)$.

[3] Again, see Broome (1995) for details.

[4] It should be easy to see, looking at matrix **1**, that there is a perfect symmetry between the A-based reasoning, and the B-based reasoning.

So $E[A|B=y] > E[B|B=y]$, for any y. In other words, no matter what the amount in envelope B is, the expected amount in A is higher, and so you ought to keep envelope A. Therefore, you reason, you ought to keep envelope A. But, you have used the same reasoning to yield contradictory demands on a rational agent. So, something must have gone wrong with that reasoning. What? This is the *paradox of the two envelopes*.

Chalmers (2002), followed by Dietrich and List (2005), argue that what the paradox shows is that the step from *switch for any value of* x to *switch simpliciter* is not justified. In other words, they take the paradox of the two envelopes to be a counterexample to the (STP_U). Indeed, the set $\mathcal{P}_\mathcal{A} = \{A = x_i\}_i$ forms a partition of Ω, and, conditional on any element of this set, the agent ought to switch to envelope B. Furthermore, the set $\mathcal{P}_\mathcal{B} = \{B = y_i\}_i$ forms a partition of Ω, and, conditional on any element of this set, the agent ought to stick with envelope A. So, the Sure-Thing Principle so formulated yields contradictory demands on a rational agent. This prompts the questions: what is wrong with this version of the Sure-Thing Principle; and can we formulate a correct version?

These questions are particularly pressing because there is no obvious other way of supplying a recommendation to the agent who is wondering whether to keep envelope A or switch to envelope B. One possibility could be to decide on the basis of the unconditional expectation values of each of the acts, $E[A]$ and $E[B]$, which are both infinite, and thus equal. But this is unappealing: as Dietrich and List (2005) point out, such a reasoning would recommend indifference between any acts X and $X+1$ defined on an infinite space, even though the latter yields an extra unit in every possible state of the world. Another possibility could be to determine whether the unconditional expected gain from switching, $E[B-A]$, is positive; but unfortunately, it is undefined in this case. Note that neither of these options would be problematic if the random variables A and B had finite expectation values.[5] This indicates that the question marks surrounding the Sure-Thing Principle are particularly pressing in infinite decision problems.

3.

The paradox of the two envelopes shows us that the the (STP_U) is false: if we allow any arbitrary partition $\mathcal{P}$ of Ω in its antecedent, the principle entails a contradiction. Dietrich and List (2005) propose to restrict the applicability of the principle. Where $\mathcal{P}$ is a *state-wise partition* of Ω just in case, for every state $w \in \Omega$, there is a proposition in $\mathcal{P}$ that picks out that state and only that state;

[5] And in fact, the paradox does not arise if the set of possible values contained in the envelopes is finite, as argued by Jackson et al. (1994).

their proposed revision is:

> **State-Restricted Sure-Thing Principle (STP$_{\text{St}}$).** Let $\mathcal{P} = \{P_1, P_2, ...\}$ form a state-wise partition of Ω. If, for every proposition $P_i \in \mathcal{P}$, an agent would perform an action T if she were to learn P_i, then the agent ought to perform T.

As Dietrich and List remark, the (STP$_{\text{St}}$) does not yield the paradoxical conclusion that the agent both ought to switch to envelope B and keep envelope A, in the initial case. Indeed, there is an element of the state-wise partition of Ω, $P_2 = \{w_2\} \in \mathcal{P}_{\text{St}}$, such that if the agent were to learn that proposition, she ought to keep envelope A. Furthermore, there is another element $P_1 = \{w_1\} \in \mathcal{P}_{\text{St}}$ such that, if the agent learnt that proposition, she ought to switch to B. So, it is not the case that an agent would switch to envelope B if she were to learn any element of $\mathcal{P}_{\text{St}}$, nor is it the case that she would stick with envelope A if she were to learn any element of $\mathcal{P}_{\text{St}}$. This entails that the (STP$_{\text{St}}$) cannot be applied in the two envelope case, and as such cannot be used to yield a recommendation to switch, nor a recommendation to keep. So, if we adopt this version of the Sure-Thing Principle, we can avoid the two-envelope paradox. Furthermore, we can retain the intuition behind Savage's house-hunting story. The agent in that situation is considering two possible states: w_D where the Democrat wins, and w_R where the Republican wins. Furthermore, if the agent were to learn either of the elements of the state-wise partition $\mathcal{P}_H = \{\{w_D\}, \{w_R\}\}$, he would decide to buy. So, the (STP$_{\text{St}}$) yields the (correct) recommendation for the agent to buy the house. So far, this speaks in favour of Dietrich and List's proposed restriction of the Sure-Thing Principle.

The (STP$_{\text{St}}$) is too restrictive, however: there are cases, relevant to the paradox of the two envelopes, where this version of the principle is silent when arguably it should yield a recommendation. The first such case, which comes from Cargile (1992),[6] I call the *coin flip case*. In that case, you are given envelope A, and you are told that it contains some amount of money from the set $\{1, 2, 4, 8, ...\}$, with the Broome probability function p presented in §2. Someone looks inside your envelope. If that person sees that the amount in your envelope is 1, he places 2 in another envelope B. If he observes any other amount in A, he flips a coin with a $3/5$ bias towards heads. If the coin lands heads, he places half that amount in B; if the coin lands tails, he places twice that amount in B. You see the entire process, except the amounts; and are given the opportunity to switch to envelope B. What should you do?

In order to determine whether you should switch to B, you reason in the following way. If the man observed that the amount in your envelope is 1, he placed 2 in envelope B, and so you should

[6] The case presented here is a slight modification of Cargile's version: he uses a uniform probability distribution, whereas I use Broome's distribution p.

switch. If he observed that envelope A contains 2, then he placed 1 in envelope B with probability $3/5$, and 4 with probability $2/5$. This means that $E[B|A = 2] = 1 \cdot 3/5 + 4 \cdot 2/5 = 11/5 > 2$. So, if envelope A contains 2, you should also switch. In fact, it can be shown that the amount in envelope B conditional on the amount in A is:[7]

$$E[B|A = x] = \begin{cases} 2x & \text{if } x = 1 \\ 11x/10 & \text{if } x > 1 \end{cases}$$

So $E[B|A = x] > E[A|A = x]$, for any x. In other words, no matter what amount the man observed in your envelope, the expected amount he placed in envelope B is higher, and you ought to switch to envelope B. Therefore, you reason, you ought to switch to envelope B. Intuitively, this reasoning seems correct in this coin flip case: you ***should*** switch to envelope B. But, this line of reasoning cannot be generated if the Sure-Thing Principle is restricted to the (STP_{St}). This is because the partition generated by the amount in A, $\mathcal{P}_{\mathcal{A}} = \{A = x_i\}_i$, is not a state-wise partition. So, the (STP_{St}) cannot be used to infer that you should switch in the coin flip case, from the fact that $E[B|A = x] > E[A|A = x]$, for any x.

The second case where the (STP_{St}) is silent when it should not be is an alternative way to reason about the initial case; and it first appears in Jackson et al. (1994). Remember, there are two envelopes in front of you, one containing twice as much as the other, but you do not know which. You are given A, and then presented with the opportunity to switch to B. You reason as follows. Suppose that the smallest of the two amounts is 1; that is, event $\{w_1, w_2\}$ obtains. Then, either A contains 1 and B contains 2, or A contains 2 and B contains 1. Given that, by assumption, these two possibilities are equally likely, and where $S = min[A, B]$, we have $E[A|S = 1] = 1 \cdot 1/2 + 2 \cdot 1/2 = 3/2$; and $E[B|S = 1]2 \cdot 1/2 + 1 \cdot 1/2 = 3/2$. In fact, more generally, we have:

$$E[A|S = z] = E[B|S = z] = z \cdot 1/2 + 2z \cdot 1/2 = 3z/2 \text{ for all } z.$$

So, were you to learn what the smallest amount is, for any such amount, you would be indifferent between A and B. So, you reason, you should be indifferent between A and B. Just as above, this line of reasoning seems fine, and its recommendation of indifference seems correct. But, if we adopt the (STP_{St}) as the correct version of the Sure-Thing Principle, this line of reasoning is blocked. This is because, in the alternative reasoning presented here, the sample space is not partitioned state-wise: the partition at hand is $\mathcal{P}_S = \{S = z_i\}_i$.

We are faced with two options. The first is to reject the recommendation of switching to B in the coin flip case, and the recommendation of indifference in the initial case; and tolerate that

---

[7] See appendix for details.

decision theory yield no recommendation in either case. This seems like a high price to pay. The second is to adopt a version of the Sure-Thing Principle which is more restrictive than the (STP_U) so as to avoid the paradox of the two-envelopes, but less restrictive than the (STP_St) so as to yield these two recommendations. In the next two sections, I propose such a version of the Sure-Thing Principle.

4.

According to an orthodox decision theory, the action an agent ought to perform in a given situation is completely determined by the options she considers, and her beliefs and desires as relating to these options. In other words, it suffices to fix a sample space of possible states, and a probability distribution and a utility distribution over those states, in order to yield a recommendation for action.[8] Calling the sample space, probability distrubution, and utility ascriptions package a ***decision situation***; and the unit of decision theory (what you need to plug in to get a recommendation) a ***decision problem***; orthodox decision theory tells us that decision situations are decision problems. In this section, I show that the initial case (§2) and the coin flip case (§3) are identical decision situations. I then take seriously the intuition that they are distinct decision problems, and examine what the implications must be for the Sure-Thing Principle. In other words, I investigate how one might modify decision theory, if one is to allow a rational agent to be indifferent between switching and sticking in the initial case, but to prefer switching to sticking in the coin flip case. In the next section (§5), I show that this leads to a resolution of the two-envelope paradox.

It should be clear that the initial case and the coin flip case are both adequately represented by matrix **1**: the sample space, probability distribution, and utility ascriptions are identical across the two cases. Indeed, the set of possible states, Ω, and the random variables A and B on that set are the same: the possible combinations of amounts in A and B are identical in the two cases. Furthermore, by assumption, the probability function p is the same. So, the two cases are one decision situation. Yet, we have strong intuitions that a rational agent ought to, or at least *may*, act differently in each case. If we want to allow for these two cases to be distinct different problems, we need to drop the claim that a decision situation is sufficient to fix a decision problem. Something else is needed. In this section, I explore what this extra ingredient must be.

The basic idea is this. If a decision situation has several features or properties, an agent might abstract away from some features and focus on others, in a context. I call these latter features

[8] These conditions are not necessary, however: one might weaken the utility distribution requirement to a preference ordering, for instance.

salient in that context. So, for instance, the salient feature of the decision situation in the coin flip case might be the amount in A. By contrast, the salient feature in the initial case might be the smallest amount in either envelope. When a feature of a situation is salient for an agent in a context, the agent groups the possible states according to this salient feature; the salient feature generates a partition of the sample space. My contention is that decision problems are individuated by both a situation and such a salient partition. In other words, they are individuated by a situation together with a way of representing or describing that situation. Before I say more about what this kind of description-relativism is, I want to talk about what it is not.

Firstly, this kind of description-relativism is not the following. Suppose that two agents, Ali and Bea, are told that a coin will be picked from an urn and tossed. Ali is told that she will get £1 if the coin lands heads and £2 if it lands tails; Bea is told she will get £1 if the coin is a pound coin and £2 if it is a two-pound coin. Both of the agents are aware that there are four ways the coin might be: pound and tails, pound and heads, two-pound and tails, two-pound and heads. Furthermore, each agent might partition this space of possibilities differently, depending on which feature of the case to which they accord importance: Ali according to heads/tails, Bea according to pound/two-pound. So, both agents are faced with the same state of affairs, but focus on different features of this state of affairs in the context of their decision. However, this is not what I mean when I talk of salient partitions. Indeed, Ali and Bea are not faced with the same decision situation: they do not ascribe the same utilities to each of the four possible states.[9] Considerations of salience as I mean them, in the context of differentiating between the initial case and the coin flip case, arise once a decision situation has been fixed. The crucial difference between the Ali/Bea cases, and the intitial/coin flip cases, is this: whereas it is irrelevant to Ali whether the coin that lands heads is a pound coin or a two-pound coin, it is not irrelevant to the agent in the coin flip case whether the amount in B is smaller or larger than the amount in A. Non-salient features of a decision situation are still relevant to the decision problem at hand.

Secondly, considerations of salience are independent from the traditional concerns about intensionality. The insight, originating with Frege (1982), is that the properties an agent believes an object to have may depend on the way in which that object is designated. For instance, a rational agent may well believe that the morning star rises in the morning, all the while not believing that the evening star rises in the morning, even if 'the morning star' and 'the evening star' refer to the

[9] In fact, instead of being represented with the same sample space and different utility functions, Ali and Bea's cases can be represented using different *sample spaces*, where Ali considers two possibilities, heads and tails; and Bea considers two distinct possibilities, pound and two-pound. In that case, we are also dealing with different decision situations.

same object—if the agent does not know that these two descriptions corefer. Credences, just like beliefs, are intensional. That is, a rational agent can have a higher credence in the proposition expressed by 'the morning star rises in the morning' than in that expressed by 'the morning star rises in the evening', despite the fact that the morning star is the evening star. Similarly, on many accounts of utility, utility also is intensional. For any chocolate, I might assign a higher utility to eating it under the description 'a piece of dark chocolate with a touch of sea salt' than under the description 'the last chocolate in the box'.

This insight has wide-ranging consequences for the individuation of decision-theoretic problems. Suppose, for illustration, that your local bookshop is running the following lottery. If you buy a ticket, you get one of two books, each with a .5 probability. The announcement for the lottery reads that one of the books is written by Ali Smith, and the other by Aurore Dupin. You know that Ali Smith was shortlisted for the Booker Prize several times, and have never heard of Aurore Dupin. You represent your decision situation in matrix **2**:

	Ali Smith	Aurore Dupin
Buy	3	-5
Don't buy	0	0

Matrix 2

	Ali Smith	George Sand
Buy	3	10
Don't buy	0	0

Matrix 3

Decision theory (as maximisation of expected utility) recommends that you do not buy the lottery ticket, and this seems like the correct recommendation. But now suppose that you have a friend who, unlike you, knows that Aurore Dupin's pen name is George Sand, but, like you, would very much like to read a book by George Sand. His decision situation is represented in matrix **3**, and yields the correct recommendation for him; namely that he should buy the lottery ticket. What has happened in this case? It is distinct from a case where you love books by Ali Smith (and therefore derive a high utility from them), while your friend hates them (and therefore derives a low utility from reading them). There, you and your friend just have different preferences. But in the case presented above, there is a sense in which you have the same preferences: if you had access to the same information about the situation (namely, if you knew that Aurore Dupin is George Sand), you would ascribe the same utilities to that state. So, what is going on is that you and your friend perceive the very same state of affairs differently; such that the utilities you plug in to an expected utility calculation are different, and decision theory recommends different actions.[10]

[10] Of course, the same point could be made with a Frege case relating to the agents' credences, instead of utilities (or both).

However, the description-relativism with which I am concerned arises once the state descriptions, credence distributions, and utility ascriptions have all been fixed. So, I am talking about an additional layer of description-relativism. One way to put the difference is thus. Frege's insight is that the way in which an agent describes an entity has an impact on which properties the agent believes that entitiy to have. The insight at play here, by contrast, is that, once it is fixed which properties an agent believes an object to have, there are some properties that might be more salient than others to the agent. Another way to put the point is: Frege cases show that decision situations are not determined only by the state of affairs at hand, or the way the world is, but also by how the agent designates entities in the situation. By contrast, the initial/coin flip cases suggest that decision problems are not determined only by decision situations.

So, if considerations of salience are distinct from typical worries about the individuation of states, and from considerations of intensionality, what are they? I think they are the same considerations as those at play in Goodman's (1983) grue case. Suppose that there are two independent features to emeralds: their colour (green or blue) and their observation-status (observed or unobserved). This entails that there are four ways that an emerald might be: green and observed, green and unobserved, blue and observed, blue and unobserved. Suppose furthermore that an agent is considering the colour of an emerald which is in fact green and observed. The agent could describe it in two ways: as green, or as grue.[11] Both of these things are true of the emerald in question, but when the agent employs one of these descriptions she focuses on some feature of the emerald, and she partitions the space of possibilities in one way rather than another—according to the full-line or to the dashed-line ways below. If, in some context, she describes the emerald as green, and partitions the space according to the rectangular way below, we say that the *salient* feature of the emerald for that agent in that context, is its greenness.

My contention is that the distinction between the initial case and the coin flip case is one of salience, too. The agent presented with the initial case is presented with the decision situation presented in matrix **1**, where the salient feature of that situation is the smallest amount in either envelope.[12] The agent presented with the coin flip case is also presented with the decision situation

[11] Where, as usual, an object is grue iff it is green and observed or blue and unobserved; and an object is bleen iff it is blue and observed or green and unobserved.

[12] Note that, if a feature's being salient corresponds to a partition's being privileged (so that greenness being salient corresponds to the possible states being partitioned according to the rectangular way above), to say that the smallest amount is salient in the two-envelope situation is equivalent to saying that the largest amount in either envelope is salient, and it is also equivalent to saying that the total amount in both envelope is salient. This is because these three features generate a single partition of Ω. Indeed, given the setup of the two-envelope case, if the smallest amount is s, the largest amount is $2s$, and the total amount

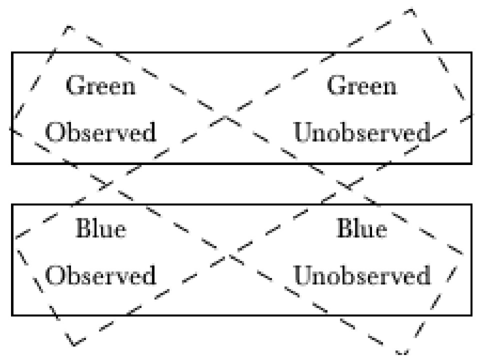

Figure 1

of matrix **1**, but this time, the salient feature of the situation is the amount in A. So, the very same possibilities are entertained in both cases, but the salient feature varies between the two cases, in such a way that the natural partitionings of possibilities differ in the two cases.

Now, this is dialectically a little too strong. My goal in this section and the next is to account for our intuition that it is rationally *permissible* for an agent to be indifferent between switching and keeping in the initial case, and to prefer switching over sticking in the coin flip case. So, I don't need to show that the salient feature is the smallest amount in the initial case, and the amount in A in the coin flip case; I merely need to enlarge the decision-theoretic framework in such a way as to allow for the possibility of a salience-based distinction between the two cases. Saying precisely when a given feature of a situation is salient to an agent in a context, that is, giving a full account of salience, is far beyond the scope of this paper. To see just how far, notice that, if I could give an account of salience, I would thereby give an account of which of an emerald's greenness or grueness is its salient feature, and why.

I can however say a few things about what makes a feature salient in a context. Firstly, it cannot be enough for the agent simply to be thinking about the feature, in order for it to become salient for that agent in that context. My thinking of an emerald's grueness does not make its grueness salient. In fact, the crux of Goodman's problem is precisely that the emerald's grueness is not salient in normal contexts—and this point could not even be uttered if thinking about a feature were sufficient to render it salient. Similarly, and this will be developed in the next section, what is wrong with the A-based reasoning in the initial case (presented in §2), is that the amount in A is not the salient feature of the situation in that case. But, if thinking about a feature sufficed to make it salient, we could not object to the A-based reasoning on these grounds.

is $3s$, in every state.

Secondly, it cannot be the case that what makes a feature salient is solely agent-independent. To see this, consider something agent-independent that might serve as the basis for distinguishing between the initial case and the coin flip case, such as the chronology with which amounts were placed in envelopes. For all we know, the amount in A was placed first, and the amount in B was determined on the basis of the amount in A and a coin flip, in both cases. In that case, the only thing that changes is how the agent represents the situations to herself. Now of course, the agent might represent situations to herself in these specific ways because of how they are presented to her. This suggests that salience may be influenced by intrapersonal factors (such as mode of presentation, language, etc.).

5.

In the previous section, I introduced the notion of *salience*. In this section, I show how it can be used to solve the paradox of the two envelopes. To do this, I point to a structural analogy between the grue paradox and the two-envelope paradox. I then use this to propose a revised version of the Sure-Thing Principle, with which the paradox can no longer be generated.

In order to generate a paradox from the introduction of the predicate *grue*, Goodman asks us to consider the application of a reasoning schema, which we can call the *straight rule*, following Jackson (1975). The straight rule is an induction rule that allows rational agents to (fallibly) infer a generalisation from a particular instance. So, for instance, it allows agents to infer that all ravens are black, from the observation of a black raven. Goodman shows that the straight rule, without any further qualifications, prescribes contradictory beliefs. Indeed, if the emerald we have been discussing is described as green, the straight rule prescribes the belief that all emeralds are green; but if the emerald is described as grue, the straight rule prescribes the belief that all emeralds are grue. These two beliefs are inconsistent: they disagree about the colour of unobserved emeralds. So, Goodman claims, the straight rule alone does not tell us what to believe about the unobserved. In order to know which such belief to adopt, we must also specify a salient description to use for the colour of the emerald.[13]

Similarly, the paradox of the two envelopes shows that a reasoning schema, the Sure-Thing Principle, can make contradictory demands on rational agents if left unrestricted. If the agent reasons using the A-based partition, the (STP_U) recommends switching to envelope B; whereas if the agent uses the B-based one, the (STP_U) recommends keeping envelope A. How can the paradox be solved then? A solution to the grue paradox must consist in restricting the applicability of the

[13] In fact, Goodman's analysis shows that this holds for any purely syntactic rule, not just the straight rule.

straight rule to salient properties. Similarly, a solution to the paradox of the two envelopes must consist in restricting the applicability of the Sure-Thing Principle to salient partitions. Once the salience restriction is in place, neither the Straight Rule nor the Sure-Thing Principle can make contradictory demands on rational agents.

If the applicability of the Sure-Thing Principle is restricted to *salient partitions*, where a salient partition is one generated by a salient feature of a decision situation, we obtain the following revision of the principle:

> **Salience-Restricted Sure-Thing Principle ($\mathbf{STP_{Sa}}$).** Let $\mathcal{P} = \{P_1, P_2, ...\}$ form a salient partition of subsets of Ω. If for every proposition P_i, an agent would perform an action T if she were to learn P_i, then the agent ought to perform T.

If we adopt the (STP_{Sa}), we can solve the paradox of the two envelopes. Remember, the paradox arose when the Sure-Thing Principle was applied to two different partitions of the sample space, $\mathcal{P}_A$ generated by the amount in A and $\mathcal{P}_B$ generated by the amount in B, and thus yielded contradictory demands on rational agents. If we adopt the (STP_{Sa}), such contradictory demands can no longer be yielded. Indeed, if only one of the amounts in A or B is salient, the (STP_{Sa}) only yields one recommendation—to switch or to stick, respectively. If both of the amounts are salient, the salient partition is a state-wise partition. To see this, consider the following. If the amount in A is salient, the salient partition of the sample space is the one where each element of the partition is associated with an amount in A. So, if both the amounts in A and in B are salient, the salient partition is one where each element of the partition is associated with an *amount in A, amount in B* pair—or in other words, with a single state. In that case, the (STP_{Sa}) does not yield any recommendation, for the reasons discussed in §3. So, the (STP_{Sa}) cannot yield contradictory demands on rational agents, and the paradox does not arise.

This is not quite enough. The (STP_{St}) was also shown not to be paradoxical, and yet was ruled as an unsatisfactory solution to the paradox. This was because, although it respected the intuition behind Savage's house-buying case, it could neither yield a recommendation in the coin-flip case, nor in the alternative reasoning about the initial case. I must therefore show that the (STP_{Sa}) performs at least as well on the house-buying case, and better in the other two.

The main charge put forward against the (STP_{St}) in §3 was that it could not allow for the alternative, smallest-amount based reasoning in the initial case; nor could it differentiate between the initial case and the coin flip case. The (STP_{Sa}) by contrast can do both of these things: it follows from the recognition that the amount in A is the salient feature of the two envelope situation in the coin flip case, whereas the smallest amount is the salient feature in the initial case. Indeed, this consideration entails that, in the initial case, the Sure-Thing Principle can be applied to the smallest-amount-generated partition but not to the A-amount-generated partition. This in turn entails that the agent ought to be indifferent between A and B, and that it's not the case that the agent ought to prefer envelope B. So, the reasoning presented in §2 is not correct, but the alternative reasoning is, in accordance with our intuitions. Furthermore, the recognition that the amount in A is salient in the coin flip case entails that the agent ought to switch to envelope B in that case. This allows us to differentiate between the initial case and the coin flip case, also in accordance with our intuitions. Again, I have not provided an argument for the claim that the smallest amount is salient in the initial case, and that the amount in A is salient in the coin flip case. But, I have shown that, unlike the (STP_{St}), the (STP_{Sa}) *can* acommodate different recommendations across the initial and the coin-flip case.

I now turn to showing that the (STP_{Sa}) yields the correct recommendation in Savage's house-buying case. Remember that Savage's agent is deciding whether to buy a house. He decides that he would if the Republican candidate wins the upcoming election, and that he would too if the Democratic candidate wins that election. Given that the Republican winning/Democrat winning feature of the situation partitions the possible states, and that this partition is salient for the agent (as Savage says, 'he considers the outcome of the next presidential election relevant"), the (STP_{Sa}) can be applied; and we get the recommendation to buy the house. So, on Savage's case, the (STP_{Sa}) performs at least as well as the (STP_{St}). In fact, it even performs better. Suppose—as is plausible—that more than one feature has a bearing on whether Savage's agent should buy a house. Maybe he cares about the reputation of the local school too, for instance. This means that he is considering four possible states: good reputation and Democrat wins, bad reputation and Democrat wins, good reputation and Republican wins, bad reputation and Republican wins. Suppose that, in one of these four states, the agent would prefer not buying a house to buying a house. In that case, the (STP_{St}) cannot yield the recommendation to buy, because there is an element of the state-wise partition where the utility of buying is strictly lower than that of not buying. The (STP_{Sa}) by contrast can still yield that recommendation—if as assumed, the political partition is salient to the agent.

The easiest way to illustrate this is to imagine that Savage's agent is a pure expected utility maximiser. Matrix **4** represents the case where the agent considers that the only feature that can impact his decision to buy a house is the outcome of the forthcoming election. Matrix **5** represents the case where another feature is also relevant, namely the reputation of the local school. It is easy to see here that the (STP_{St}) only yields a recommendation to switch in the former case, and not in the latter. By contrast, (STP_{Sa}) can be used to yield a recommendation in both cases, on the assumption that the outcome of the election is salient for the agent.

	Democrat wins	Republican wins
Buy a house	3	1
Do not buy a house	2	0

Matrix 4

	Good reputation Democrat wins	Bad reputation Democrat wins	Good reputation Republican wins	Bad reputation Republican wins
Buy a house	4	2	3	-1
Do not buy a house	2	2	0	0

Matrix 5

Of course, the failure of (STP_{St}) in this case is not particularly worrying: since we have assumed the agent to be an expected utility maximiser, we can simply use a utility calculation to yield the recommendation to buy in both cases. However, the EU-maximisation assumption here only serves to fix the preference orderings, and the same point can be made (albeit less concisely) without it.

7.

What can we take away from this discussion? We are faced with two options. Firstly, one could accept the claims I have been making throughout the paper; and in particular that the correct formulation of the Sure-Thing Principle must contain a salience-based restriction on which partitions are allowed. This implies that a kind of description-sensitivity is central to the applicability of the Sure-Thing Principle, which is distinct from and additional to the designation-sensitivity that arises from the intensionality of credences and utilities. More generally, it implies that the individuation of decision theoretic problems does not depend solely on decision situations, but also on which feature of such decision situations are salient. Fixing a sample space of possibilities, a probability distribution, and utility functions on this space are not sufficient to yield a recommendation for action, on my view: a privileged partition, generated by the salient feature of the situation for the agent, must also be specified.

Alternatively, one could reject my view, and maintain instead the commonplace view that fixing a decision situation is sufficient for obtaining a recommendation for action. This comes at a high cost; namely, violating strong intuitions about which actions are rationally permissible in the two-envelope initial and coin flip cases. We have seen that the (STP_U) yields contradictions, and that the (STP_{St}) cannot yield recommendations in the initial or coin flip cases. We might nonetheless be able to formulate a different decision rule that allows us to get a recommendation for action, and so rejecting my view does not entail silence on these cases. However, it does entail that rational agents must act in the same way in the initial case and in the coin flip case. So, ultimately, a choice must be made between a revisionary view about the individuation of decision theoretic problems, and what I think is an implausible constraint on rational agents.

References

Broome, J. (1995). The Two-Envelope Paradox. *Analysis 55*(1), 6–11.

Cargile, J. (1992). On a Problem About Probability and Decision. *Analysis 52*(4), 211–216.

Chalmers, D. J. (2002). The St. Petersburg Two-Envelope Paradox. *Analysis 62*(2), 155–157.

Dietrich, F. and C. List (2005). The Two-Envelope Paradox: An Axiomatic Approach. *Mind 114*(454), 239–248.

Frege, G. (1982). On Sense and Reference. In P. Geach and M. Black (Eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford University Press. 1980.

Goodman, N. (1983). *Fact, Fiction, and Forecast*. Harvard University Press.

Jackson, F. (1975). Grue. *Journal of Philosophy 72*(5), 113–131.

Jackson, F., P. Menzies, and G. Oppy (1994). The Two Envelope "Paradox". *Analysis 54*(1), 43–45.

Savage, L. J. (1954). *The Foundations of Statistics*. Wiley.