Bets on Hats - On Dutch Books Against Groups, Degrees of Belief as Betting Rates, and Group-Reflection

Rabinowicz, Wlodek; Luc, Bovens

Published in:
Episteme

2011

BETS ON HATS

On Dutch Books Against Groups, Degrees of Belief as Betting Rates, and Group-Reflection

Luc Bovens and Wlodek Rabinowicz

Our *Story of the Hats* is a puzzle in social epistemology. Our earlier presentations of this puzzle were rather technical. (Bovens & Rabinowicz 2009, 2010) We would like to take this opportunity to spell out an informal version of the puzzle and its solution and to explore their philosophical relevance.

It all started when one of us, Luc Bovens, read a *New York Times* article on a fascinating mathematical problem, named 'the Hat Puzzle', formulated by Todd Ebert in his doctoral thesis in 1998 (Robinson 2001). In this problem, there are *n* players, each with a hat on his head. Each hat is either black or white, with the alternatives being equally probable. This is common knowledge among the players. Each player can see the hats of the others, but not his own. Based on this information, each player must independently and privately guess the color of his own hat. Abstaining from guessing is allowed. The group will receive a prize if (i) at least one player makes a correct guess and (ii) no incorrect guesses are made. The question Ebert raised was: How should the players act in order to maximize the chance of winning the prize?

The problem is easy to solve for three players. The strategy that maximizes the probability of winning the prize for the group is the following: If you see two hats of the same color, then guess that your own hat is of the opposite color; if you see two hats of different colors, then keep quiet. There are eight possible combinations, all equiprobable, viz. (1: White; 2: White; 3: White), (1: White; 2: White; 3: Black), etc. Of these eight combinations, there are six in which not all the hats are of the same color. The strategy leads to a win in each of the six combinations in which the hats are not of the same color and a loss in each of the two combinations in which the hats are of the same color. So the strategy has chance ¾ of being a winning strategy.

More generally, Ebert's question has been answered for all n such that for some integer k, $n = 2^k - 1$. The chance of winning can then be maximized to $n/(n+1)$. But a solution for an arbitrary number of players is still unavailable, as far as we know.

The other one of us, Wlodek Rabinowicz, thought one could use the Ebert scenario to set up a new paradox of group rationality. In the Prisoner's Dilemma, individual rationality adds up to collective irrationality, due to the fact that each player aims to promote her own goals rather than common goals. Now the aim was to use the Hats puzzle to devise a Dutch Book showing that individual rationality could add up to collective irrationality even if the players aim to promote common goals. In more detail, the aim was to show that a clever bookie could exploit a group of agents who (i) are Bayesians, (ii) have common priors, (iii) have common goals, and (iv) have common knowledge of (i), (ii) and (iii), if such agents are exposed to different information and are expected to make decisions independently. This we have named 'the Story of the Hats'. Situations in which this happens involve violations of what might be called the Group-Reflection Principle. The Dutch book that can be set up against groups that violate Group-Reflection is a version of van Fraassen's well-known Dutch book against individual agents who violate the standard individual version of Reflection.

Unfortunately, as it turned out, the argument leading to the paradox is flawed, but it is flawed in interesting ways. It was based on the betting interpretation of the subjective probabilities, but ignored the fact that this interpretation does not take into account the strategic, i.e. game-theoretic, considerations that might influence an agent's betting behavior. If such considerations are taken into account, as they clearly should, then – as we are going to show – the Dutch book construction crumbles.[1] This destruction process has been pursued by both of us together.

Thus, what the Story of the Hats helps us see is different from what originally appeared to be its attraction. The argument that individual rationality in pursuing a common objective can add up to collective irrationality does not go through after all. Instead, the lesson to be learned concerns the interpretation of probabilities in terms of fair bets and, more generally, the role of game-theoretic considerations in epistemic contexts. Another lesson concerns Group-Reflection, which in its unrestricted form is a highly counter-intuitive principle. We consider how this principle of social epistemology should be formulated so as to make it tenable.

1. *A Dutch Book?*

Here is how the Story of the Hats goes. Suppose that a group consisting of three persons learns the following: Each of them will be given a hat to put on in the dark. The hat's color, either black or white, will be decided in secret and independently for each individual by a toss of a fair coin (say, she gets a black hat if tails and a white hat if heads). Then the lights will be turned on and each person will be able to see the hats of the other two members of the group, but not yet her own. All of the above is common knowledge in the group.

We distinguish three temporal stages: Stage 1, in the dark; stage 2, when the lights are turned on; stage 3, the aftermath, when the color of each hat is revealed to everyone.

Consider the following proposition:

(*A*) Not all hats are of the same color.

At stage 1, given what the group members know at that stage, their probability for *A* is ¾. For, as we have seen in the introduction, the chance of *A* being true is then known to be 6/8 (= ¾), and according to David Lewis' principal principle, a (Bayesian) rational agent will let her subjective probabilities be determined by her knowledge of chances. (Cf. Lewis 1980.)

At stage 2, however, the group members' probabilities for *A* may diverge. (i) They *will* diverge if *A* is true. If *A* is true, there will be two hats of the same color and one hat of a different color. Consequently, at stage 2, one person, *j*, will be seeing two hats of the same color, while the other two persons will be seeing one black hat and one white hat. The probabilities of the latter two for *A* will therefore be 1, since they can infer from what they see that *A* is true. Person *j*' s probability for *A* will be ½: *A* is true just in case the color of her hat differs from that of the hats of the other two players and *j* knows that the prospect of her hat being of the same color is just as likely as the prospect of it being of a different color, since she knows that the color of her hat was chosen independently and at random. (ii) If *A* is false, on the other hand, all three persons will be seeing two hats of the same color, which means that each of them will assign probability ½ to *A*.

We assume that the three persons in the group have common assets with which they manage their bets and a common objective – viz. to maximize the group's holdings – but for the

---

[1] We are indebted to Anthony Williams for alerting us to this problem.

argument to follow it is crucial that each person can independently draw on the common assets of the group. Thus, they are a group with common assets and with common interests, but with decentralized powers: each member of the group is an independent decision-maker.

We first sketch the betting interpretation of subjective probabilities (or degrees of belief or credences, as we shall also call them). A bet on a proposition X has a price C and a non-zero stake S, where the latter is the monetary prize to be won if one buys the bet and X turns out to be true. It is said to be *fair* for a given agent if and only if the latter is willing to take each side of the bet, i.e., if she is just as willing to buy the bet as to sell it. Assuming that there is such a bet on a proposition X and that the C/S ratio is constant for different fair bets on X (i.e., that changing the stake requires a proportional change in the price), then this ratio is the agent's *betting rate* for X: $P(X) = C/S$. These assumptions about the existence of fair bets and the constancy of the C/S ratio are controversial, but we shall ignore this in what follows. On the betting interpretation, the agent's degrees of belief for various propositions are her betting rates for the propositions in question. Note that this makes the expected monetary value of a fair bet equal to zero: If the agent's degree of belief $P(X)$ equals the betting rate C/S, then $(S \times P(X)) - C = (S \times C/S) - C = 0$. This also explains why the agent is equally willing to buy such a bet as to sell it: Gains for the seller become losses for the buyer and vice versa; as a buyer, the agent's expected monetary value is $(S \times P(X)) - C$, as a seller, her expected monetary value is $C - (S \times P(X))$; if one of these differences equals 0, then so does the other, and hence the expected monetary values of buying and selling are identical.

The identification of degrees of belief with betting rates has the advantage of making degrees of belief observable and measurable. It also makes it possible to give pragmatic arguments for various epistemic rationality constraints on beliefs such as the requirements that degrees of belief obey standard probability axioms, conditionalization, reflection, etc.: An agent whose degrees of belief do not satisfy these requirements will have betting rates that make her vulnerable to exploitation.

What we are after here is a betting arrangement that is meant to exploit the group as a whole, despite the fact that each member's degrees of belief satisfy the constraints of individual rationality and despite the fact that all of them start out with the same degrees of belief (the same priors). Here is how such a Dutch book could be set up. Suppose that at stage 1, i.e., before the lights are turned on, the bookie offers to sell a single bet on the proposition A with a stake of \$4 at a price of \$3, and subsequently, after the lights are turned on at stage 2, he offers to buy a single bet on *A* with a stake of $4, as before, but at a lower price of $2. Suppose that all of the above from the outset is common knowledge among the players.

Since at stage 1 each group member assigns probability ¾ to *A*, each should be willing to buy the first bet, whose price-stake ratio is $3/$4. And since at stage 2 there will be at least one agent who assigns probability ½ to *A*, there will be at least one agent who should be willing to sell the second bet, whose price-stake ratio is $2/$4. (If there are several volunteers, we can assume that the bookie will pick one of them at random.) Since (i) the bets are on the same proposition, (ii) the stakes are equal, and (iii) the price of the second bet is lower, the bookie can make a Dutch book – whether all hats are of the same color or not, he has a guaranteed profit of $1.

More precisely, consider what happens at stage 3, when the colors of the hats are revealed to everyone. If *A* turns out to be true, the bookie pays out the stake of $4 on the first bet and collects the same amount on the second bet. If *A* turns out to be false, no stake-payments will be made. But, whether *A* is true or not, the bookie makes the net profit of $3 - $2 = $1 on the price difference: First he sells a bet on *A* and then buys it back at a lower price. The group has thus been exploited by a bookie without any superior knowledge, even though all of the group members start with the same information and each of them has updated her probabilities in a perfectly reasonable way and acts on these probabilities in a seemingly perfectly rational manner.[2]

2. Group-Reflection

The well-known *Reflection Principle* requires an agent to adjust her present probabilities to her expected future probability assignments. In one of its versions, the principle may be formulated as follows:

---

[2] If you find the diachronic form of this argument worrying, consider the following *synchronic* version of the Story of the Hats, this time with four agents instead of three. Three of the agents, Alice, Barbara and Carol, have been equipped with hats. Each sees the hats the other two agents wear, but not her own. The fourth agent, Diana, is hatless and doesn't see the hats of the other three women. The color of each hat, either black or white, has been chosen in secret, independently and at random. All this is common knowledge in the group. Now Diana's probability for the proposition *A* is ¾. At least one agent in the group, Alice, Barbara or Carol, ascribes probability ½ to *A*. All this is common knowledge. Thus, a clever bookie who doesn't know more than what's common knowledge in the group can set up a synchronic Dutch book against the group as a whole: He can sell to Diana a bet on *A* at her odds (price 3, stake 4) and at the same time offer to buy a single bet on *A* with the same stake and a lower price (price 2, stake 4), knowing that that this second bet is fair for at least one of the other three agents. So, again, he will profit due to the price difference.

If $P(P'(X) \leq k) > 0$, then $P(X \mid P'(X) \leq k) \leq k$.

Here X is an arbitrary proposition, P stands for the agent's probability at a time point t, and P' for her probability at some point t' that is not earlier than t. So, what the principle says is the following. Assume that one deems it possible that one's future probability for X is less than or equal to k. Then, conditionally on this possibility being the case, one should *now* assign to X a probability less than or equal to k. One's current probability, conditional on a hypothetical future probability assignment, should reflect this future assignment. This Reflection Principle thus demands full trust in one's future self.

As is well known, Bas van Fraassen (1984) has shown that agents who violate Reflection are vulnerable to diachronic Dutch books.

Now one might ask: If we want to make a *group* invulnerable to a Dutch Book, then what sort of reflection principle should the group abide by? Consider the following principle:

The Group-Reflection Principle:

For any i in G, if $P_i(\exists j \in G: P'_j(X) \leq k) > 0$, then $P_i(X \mid \exists j \in G: P'_j(X) \leq k) \leq k$.

Here, P_i stands for i's probability at a time t, while P'_j for j's probability at some time t' that is not earlier than t. What the principle says is the following. Assume that you deem it possible that some group member's future probability for X won't exceed k. Then, conditional on this possibility being the case, each group member should now assign to X a probability not exceeding k. As it stands, this principle is quite counterintuitive, even if the group members are known to be fully epistemically rational. We will discuss the principle's intuitive standing in section 5.

In our Story of the Hats, the Group-Reflection Principle is violated, as is easy to see. At stage 1, each individual i in the group G that consists of three persons is certain that at stage 2 there will be some person j in the group whose probability for the proposition A will be ½. Then it is easy to show that, by Group-Reflection, the players i at stage 1, should set $P_i(A) \leq ½$,[3] but they violate Group-Reflection, since they set $P_i(A) = ¾$.

---

[3] Let P and P' stand for probability assignments at stages 1 and 2 respectively. Then, by construction, for all i, j in the group, (i) $P_i(\exists j \in G: P'_j(A) = ½) = 1$; Hence, (ii) $P_i(\exists j \in G: P'_j(A) \leq ½) > 0$, and so, by Group-Reflection, (iii) $P_i(A \mid \exists j \in G: P'_j(A) \leq ½) \leq ½$; From (i), also (iv) $P_i(\exists j \in G: P'_j(A) \leq ½) = 1$; And so, from (iii) and (iv), by the probability calculus, $P_i(A) \leq ½$.

As we have argued, the group in our example seems to be vulnerable to a diachronic Dutch book. In fact, similarly vulnerable is *any* group that violates the Group-Reflection Principle. To set up a Dutch book for such a group, we make appropriate adjustments in van Fraassen's construction. Here is how this can be done.

Let t be any time point and G any group of agents. Let P_i specify agent *i's* probability assignments at t and P'_i specify i's probability assignments at a time t' that is not earlier than t. Suppose that the Group-Reflection Principle is violated:

(*VIOL*) For some i in G, $P_i(E) \geq z$ and $P_i(X \mid E) \geq y$

with $E = \exists j \in G\colon P'_j(X) \leq k$,

where $z > 0$ and $y > k$.

We let the bookie make an offer to sell to the first-comer in G the following conditional bet at t:

Bet 1: Bet on X, conditional on E, with a positive stake S and a price $C = S \times y$.

That the bet is conditional means that it will be called off if the condition E will turn out not to be fulfilled. As is easy to see, the bookie's offer is fair or more than fair: The group member i will be prepared to buy Bet 1 since the price-stake ratio of that bet isn't higher than her conditional probability for X given E: $P_i(X/E) \geq y = (S \times y)/S$. Person i's expected winnings from that bet if E is the case, $S \times P_i(X/E)$, are at least as large as the price of the bet, $S \times y$.

To offer the conditional bet at the right odds, the bookie needs to know that the Group-Reflection Principle is violated and how it is violated: He needs to know, for some $y > k$, that the group contains a member who assigns a probability of at least y to X given E, thereby violating Group-Reflection. To guarantee that the bookie does not know more than any member of the group, we need to assume that *everyone* in G knows as much. The argument to follow is meant to show that, if the above is the case, a bookie who lacks superior knowledge can set up a Dutch book against G. More precisely, without any need of superior knowledge on the part of the bookie, a Dutch book can be set up against a group G if (*VIOL*) and every j in G knows that (*VIOL*).[4]

[4] A similar remark applies to the violations of Reflection by a single agent. A Dutch book without superior knowledge on the part of the bookie can be set up against such an agent only if she herself knows her own probability assignments that violate the principle. This self-introspection on the part of the agent is usually tacitly assumed in the discussion of the Reflection Principle.

Let us continue our description of the Dutch book. At the second stage, at t', the bookie offers to *buy* the following unconditional bet from the first-comer in G:

Bet 2: Bet on X, with a stake S and a price $S\times k$.

If that offer is accepted, the bookie pays $S\times k$ to the seller and then collects S iff X is true.

There are two possible cases.

(i) E is false. Then for every j in G, j's probability for X at t' exceeds k, which means that no j will be prepared to sell Bet 2 to the bookie. For j's expected value from selling Bet 2 is $(S\times k) - (S\times P'_j(X))$, which is negative if $P'_j(X) > k$.

(ii) E is true. Then there will be some j in G such that $P'_j(\mathrm{A}) \leq k$. There is thus a willing seller of Bet 2, since j's expected value from selling that bet is non-negative. (In case there are several persons willing to sell Bet 2, the bookie will pick one of them at random, or buy that bet from the first-comer.)

We conclude that if E is true, Bet 1 will be in force and Bet 2 will find a seller, j. The bookie will pay the price $S\times k$ for Bet 2 and he will receive the price $S\times y$ for Bet 1. Since $y > k$, the bookie's total price balance, $S\times(y\text{-}k)$, will be positive. As for the stake S that the bookie has to pay out to i on Bet 1 in case X is true, he will receive exactly the same amount in that case from j on Bet 2, which means that the bookie's net gain, whether X is true or not, will be positive: $S\times(y\text{-}k)$.

Still, as things stand, the bookie won't make any profit if E is false:[5] For then the first bet will be off and the second will not be accepted. To cover this eventuality, let us modify the set-up: Suppose that the bookie at t offers to sell *two* bets: the conditional Bet 1 and a *side bet* on E at a stake $S\times(y\text{-}k)$ and a price $z\times S\times(y\text{-}k)$. Since i's probability for E is at least as high as z, the offered side bet is fair or more than fair for i, which means that she is willing to take on this side bet as well. Now since z is positive and $y > k$, the price for the side bet is positive. Thus, if E is false, the bookie still gains: Bet 1 will be called off, but the bookie will have the amount $z\times S\times(y\text{-}k)$ which he received as the price for the side bet. On the other hand, if E is true,then the bookie's overall profit will be $S\times(y\text{-}k)$ [= the net gain from Bets 1 and 2] - ($S\times(y$-

[5] In our Story of the Hats, we have excluded this possibility. In that story, $E = \exists j\in G$: $P'_j(A) \leq ½$ holds by the construction of the example. In other words, that at stage 2 some agent will assign probability ½ to A is something that is bound to happen.

k) - $z \times S \times (y-k)$) [= the net loss from the side bet on E] = $z \times S \times (y-k)$.Thus, the bookie will profit no matter what happens.

3. Probability divergence and epistemic idealizations

Let us return to the example we have started with. One reaction might be that it isn't at all strange or paradoxical that groups with common assets and decentralized powers can be exploited if their members have divergent probability assignments. Thus, to illustrate, think of a husband and wife who assign different probabilities to some proposition *X*. Suppose, for example, that the wife's probability for *X* is lower than her husband's. Then a bookie can sell a bet on *X* to the husband and buy that same bet from the wife for a lower price, thereby making a sure profit. If the husband and his wife know each other's probabilities, then the bookie needn't know more than each of the spouses in order to make a sure profit. Why bother then with the Story of the Hats? Why is the husband-and-wife example not sufficient to make our point about group vulnerability?

One might think that the latter example is less compelling since, for the exploitation to become possible, it presupposes that the group members' probability assignments actually diverge. In the Story of the Hats, on the other hand, it is not *necessary* that the probabilities of the group members should diverge at any stage. At stage 1, their (relevant) probabilities are the same. And at stage 2, they will also be the same *if* all hats happen to be of the same color. Divergence at stage 2 is possible but not necessary. It is the possibility of divergence rather than its actual occurrence that is essential for the example in question.

However, even the husband-and-wife case can be modified in so as to make actual divergence unnecessary. Such a modified version is presented in Christensen (1991, see esp. pp. 244ff). Essentially, as Christensen points out, it is enough if, say, the husband considers it possible that his wife's probability assignments diverge from his own but is not prepared to adjust his conditional probabilities accordingly. If the husband thinks that his wife tends to be unduly pessimistic (or unduly optimistic), then his conditional probabilities will violate the Group-Reflection Principle[6] and thus, as we showed in section 2, the exploitation will be possible,

[6] If the husband thinks the wife is unduly pessimistic with respect to a proposition *X*, then it may well happen that his conditional probability $P_{\text{Husband}}(X \mid P_{\text{Wife}}(X) \leq k) > k$, in violation of Group-Reflection.

quite independently of whether the wife's probability assignments actually diverge from her husband's or not.

The more deep going difference between the two examples is the following: In the husband-and-wife story, it appears that hat something of an epistemic nature is amiss and that this is what brings about the couple's divergent probabilities or the suspicion that these probabilities might diverge. It is not made explicit what it is that is amiss. Maybe one of the parties is unduly optimistic or pessimistic in their reasoning. Or maybe they start from different priors. Or, in Christensen's modified version, it is sufficient that one of the parties believes that the other is unduly optimistic or pessimistic or starts from different priors. In the Story of the Hats, on the other hand, there is nothing epistemically amiss about the parties – they start from the same priors, they do not process information in an unduly optimistic or pessimistic manner, and they have common knowledge hereof. The bookie is not exploiting some kind of epistemic shortcoming. And yet, they are still vulnerable, as a group, to a Dutch book. Or, at least, so it seems.

4. The Dutch book deconstructed

So, seemingly, in the Story of the Hats, the bookie has succeeded in making a Dutch book. This seems worrisome. A Dutch Book is a mark of *ir*rationality. There seems to be some breakdown of group rationality, as in Christensen's cases. One can also juxtapose the Story of the Hats to the Prisoner's Dilemma. In a Dutch Book, the group loses. In the equilibrium solution to the Prisoner's Dilemma, the group does worse than it could have done. Now this is due to the fact that each player acts to her own advantage. But in the Story of the Hats, the players act in the group interest (unlike in the Prisoner's Dilemma) and they do so in a fully rational manner (unlike in Christensen's husband-and-wife cases).

Fortunately, the Dutch book we have presented is spurious. (Or unfortunately? It's always a pity when a nice paradox goes down the drain.) Showing this will hopefully prove instructive.

Let's focus on the second bet on proposition *A*, the one that the bookie offers to buy at stage 2. The price he offers to pay is \$2 and the stake of the bet is \$4. Thus, the price-stake ratio is ½. As we know, at stage 2, there is at least one person in the group who sees two hats of the same color and therefore assigns probability ½ to *A*. We have been assuming that for every person in this position the bet offered by the bookie is fair, which implies that anyone in this position should be willing to sell such a bet to the bookie. But is this really the case? Suppose that Alice is a person in this position and consider how she might reason about the bookie's offer.

There are two possible cases: either A is true or A is false.

If A is true, i.e., if not all hats are of the same color, then Alice is the *only* person in the group who sees two hats of the same color. The other two group members see two hats of different colors, which means they know that A is true. So they will certainly ignore the bookie's bid to buy a bet on A. This means that, if Alice declares herself willing to sell the bet to the bookie, she will be the only volunteer and her offer will be accepted. But if she does sell the bet on A, she will cause a net loss to the group: she will collect the price of $2, but then will have to pay out the stake of $4 to the bookie. If she abstained from declaring herself willing, no other group member would come forward and the loss to the group would be avoided.

If A is false, on the other hand, selling the bet on A to the bookie would lead to a net gain for the group. But if A is false, then the other two group members are in the same situation as Alice: each also sees two hats of the same color and therefore assigns probability ½ to A. Alice might think, therefore, that in this situation it's fine to abstain from declaring her willingness to sell the bet, since there are other group members who could declare themselves willing to sell the bet. In other words, she might think that in such a situation it's fine to let others do what needs to be done.

So, Alice might draw the conclusion that she should abstain from declaring herself willing to sell the bet, because the declaration of willingness would be positively harmful if A is true and abstaining from it does not lead to any loss of opportunity for the group if A is false.

There is something fishy, though, about this reasoning. If A is false, then every group member is in the same position as Alice. This means that, if her reasoning is sound, then her conclusion would be drawn by the other two group members as well: neither one of them would declare herself willing to sell the bet. This undermines the second part of Alice's argument: if A is false, then her decision not to come forward *would* involve an opportunity loss for the group.

Still, even if the reasoning is faulty, we can see that the standard betting interpretation of subjective probabilities is too coarse to deal with the situations in which an agent makes decisions on bets *while interacting with other agents*. There are various strategic considerations that become relevant in situations of this kind, which complicates matters.

An appropriate tool for the study of interactions is game theory. Let us therefore consider the situation created by the bookie's bet offer from a game-theoretic perspective. There are three players – the three members of the group. Each player *i* can either volunteer to sell the bet to the bookie or abstain from volunteering. Since she can also opt for a mixed action as well, we can describe her options as the set of probability values: she can volunteer with a certain probability *p* between 0 and 1, and abstain with probability 1-*p*. What a player decides to do depends on her information at stage 2: She then either sees two hats of the same color or two hats of different colors. *i*'s strategy can therefore be described as a pair of probabilities (*p*, *q*), where *p* specifies the probability of volunteering if she sees two hats of the same color, while *q* the probability of volunteering if she sees two hats of different colors. Each strategy profile $< (p_a, q_a), (p_b, q_b), (p_c, q_c) >$ for players Alice, Barbara and Carol yields payoffs that are identical for all players, since the players' sole objective is to increase the common resource pool.

Now, since the bet on A the bookie wants to buy is offered at odds that correspond to the probability for A of a player who sees two hats of the same color, and since a player who sees two hats of different colors knows that the bet will be lost by the group, it might be thought that the following profile provides a solution for this game: < (1,0), (1,0), (1,0) >. I.e., each player volunteers to sell the bet if she sees two hats of the same color, but stays back otherwise. However, a moment's reflection shows that this cannot be the right solution, since the profile in question is *not* a Nash equilibrium, i.e., there is a reason for at least one of the players to deviate from it provided that the other two players do not deviate. To see that, let us look at the matter from the point of view of the player who sees two hats of the same color. There will be at least one such player at stage 2 and let's suppose it will be Alice. The outcome of < (1,0), (1,0), (1,0) > is that the bet will be placed. In terms of Alice's credences at stage 2, the expected payoff of the bet for the group is 0. But suppose that she were to unilaterally deviate to < (0,0), (1,0), (1,0) >. The expected payoff for the group would then be positive. That all the hats are of the same color (and the group wins $2) is just as probable from Alice's perspective as that they are not (and the group loses $2). If they are of the same color, then Barbara and Carol will step forward if < (0,0), (1,0), (1,0) > is played and the bet will be sold. If they aren't of the same color, then Barbara and Carol will not step forward and the bet won't be sold. Hence the expected payoff of < (0,0), (1,0), (1,0) > from Alice's perspective equals $1—there is a 50% chance of winning $2 and a 50% chance of the group refraining from betting. So < (1,0), (1,0), (1,0) > is not a Nash equilibrium since a unilateral deviation by Alice increases the payoff of the group.

Somewhat surprisingly, however, it turns out that < (1,1), (1,1), (1,1) > *is* a Nash equilibrium. On this profile, every player volunteers to sell the bet *whatever* information she receives at stage 2. Consider the situation from Alice's point of view: Whether she sees two hats of the same color or of different colors, she knows that the other two players will volunteer whatever she does. So it doesn't matter what she does herself: the bet will be accepted by the group anyway. Consequently, there is no reason for her to unilaterally deviate from the profile under consideration.

Does this show that the bookie can rest assured that he will manage to buy the bet at stage 2 and in this way complete his Dutch book? No, it doesn't. Apart from the fact that < (1,1), (1,1), (1,1) > is not a *unique* Nash equilibrium in this game (which means there is no guarantee that players will settle on that one rather than on some other solution), it is easy to show that this equilibrium has a serious disadvantage: it's not *trembling-hand perfect*.[7] I.e., as soon as we entertain the possibility that players have a slight tendency to 'tremble', i.e. that they might fail to move as the strategy profile requires, volunteering to sell the bet when you see two hats of different colors (i.e. when you know that selling the bet will cause a loss to the group) becomes plain stupid. After all, in such a situation you don't gain anything by volunteering and you can cause a positive loss if both other players tremble and abstain from volunteering to sell the bet.

Note that this argument applies not only to < (1,1), (1,1), (1,1) >, but to every profile in which at least one of the players volunteers with some positive probability if she sees two hats of different colors. And since we already know that < (1,0), (1,0), (1,0) > is not even a Nash-equilibrium, it follows that there is no strategy profile < (1, q_a), (1, q_b), (1, q_c) > that is a trembling-hand perfect Nash equilibrium. This means that the bookie cannot rest assured that the player who sees two hats of the same color will step forward and sell the bet to him. He knows that there will be at least one such player, but there is no guarantee that any of them will be stepping forward. Consequently, the Dutch book crumbles.

There are several trembling-hand perfect Nash equilibria in this game, but one of them is especially salient because of its symmetry properties: <(0,0), (0,0), (0,0) >. That this profile is

---

[7] The idea of a trembling-hand perfect equilibrium goes back to Selten (1975).

a Nash equilibrium is clear, since from the point of view of the player who sees two hats of the same color, the bet offered by the bookie has a zero payoff, which means it doesn't matter whether the bet will be made or not. While if a player sees two hats of different colors, he knows that the bet would be disadvantageous, which means she never has any reason to deviate from her resolution to stay back, whatever she believes about other players. Thus, in both cases, nothing is gained by a unilateral deviation from the profile that requires each player to stay back whatever information she gets.

To show that <(0,0), (0,0), (0,0) > is trembling-hand perfect, we only need to consider it from the point of view of a player, say, Alice, who sees two hats of the same color. Suppose Barbara and Carol have a slight tendency to 'tremble' – each of them might step forward with some small probability, where that probability does not depend on the information that they receive. There are two possibilities, each with probability ½: Barbara and Carol see (i) two hats of the same color, or (ii) two hats of different colors. In case (i), if someone volunteers, the expected payoff to the group is 2; while in case (ii), it is –2. In no one volunteers, the expected payoff is 0. In each of those cases, there is some positive probability δ that Barbara or Carol (or both) will volunteer because of their trembling hands.[8] So, if Alice stays back, the expected outcome of her action is $\frac{1}{2}(2\delta) + \frac{1}{2}(-2\delta) = 0$. And obviously, if Alice volunteers, the bet will be sold to the bookie, so the expected outcome of volunteering will again equal zero: $\frac{1}{2}(2) + \frac{1}{2}(-2) = 0$. Consequently, she has no reason for a unilateral deviation from her resolution to stay back.[9]

5. *Dutch books and Group-Reflection*

What are the implications of this deconstruction of the Dutch book in the Story of the Hats for the general Dutch book argument for the Group-Reflection Principle? The Dutch book that we laid out in the section on Group-Reflection is an exploitation set-up that is meant to be applicable whenever Group-Reflection is violated. Does the deconstruction of this Dutch book in the case of the Story of the Hats show that the exploitation set-up in question is

---

[8] If the probability of tremble equals ε for each player, then $\delta = 1 - (1 - \varepsilon)^2 = 2\varepsilon - \varepsilon^2$.

[9] Note that the same argument can be employed in a proof that a non-symmetric profile <(1,0), (00), (0,0) > is also a trembling-hand perfect Nash equilibrium. We are indebted for this point and for other useful comments to Chlump Chatkupt, who has helped us to see that our previous formulation of the game-theoretic approach to the Story of the Hats in Bovens&Rabinowicz (2010) wasn't satisfactory.

generally faulty and unworkable? It does not. It all depends on whether the game-theoretic considerations that we have adduced to undermine the Dutch book in the Story of the Hats are applicable in other cases in which Group-Reflection is violated. A crucial step in the Dutch book against violators of Group-Reflection is this one (we quote): "At the second stage, at *t'*, the bookmaker offers to *buy* the following unconditional bet from the first-comer in *G*:

Bet 2 on *X*, with a stake *S* and a price $S \times k$.

If that offer is accepted, the bookie pays $S \times k$ to the seller and then collects *S* iff *X* is true."

The claim was that if *E* is true, i.e., if there is some *j* in *G* who at *t'* assigns to *X* probability lower than or equal to *k*, then the bookie will find at least one group member who will be willing to sell Bet 2 at *t'*. The bookie is supposed to be able to find such a person *j* because, from the point of view of at least one *j*, Bet 2 has non-negative expected value in terms of her probabilities for *X*. Now, this step is problematic if there may be *several* such *j*'s at that stage, on the assumption that *E* is true. Then there are strategic considerations that such agents *j* need to take into account—and this was precisely the predicament of the group members at stage 2 in the Story of the Hats. However, if there can be only *one* such *j* on the assumption that *E* is true, then game-theoretic considerations do not apply and the Dutch book is unassailable—as in Christensen's husband-and-wife story.

As a general epistemic principle, Group-Reflection is highly counter-intuitive. As it stands, this principle requires us to respect the probability assignments of every other group member, even those who have different priors or whom we do not consider to be fully rational, epistemically speaking (cf. the case of husband-and-wife mentioned above). Also, the principle requires us to respect the probability estimates of group members who might have less evidence at their disposal than we do. Surely, this is not right. In the case of individual Reflection, we can assume that our future self will have access to all that we know and possibly to some new evidence. This explains why there is a good reason to respect that future self's probability estimates (as long as we do not have good reason to expect that we will become epistemically corrupted in the future[10]). But even if we modify Group-Reflection along these lines and restrict the scope of the principle to group members who have the same priors, are fully epistemically rational and have all the evidence that we have and possibly more, the so restricted principle can still lead us astray, as we have seen in the Story of the

[10] For a discussion of (individual) Reflection and counterexamples to Reflection, see Bovens (1995).

Hats. There, the agents at stage 1 are certain that at stage 2 one of them will on good epistemic grounds have probability ½ for *A* and that this person will know more than they know. But they still ascribe probability ¾ to *A* and are perfectly justified in doing so.

To get a tenable version of Group-Reflection, we need to restrict the principle even further and require that the group members we can rely on should have at least as much information as every other member in the group. Here is a tentative formulation that one might consider:

The Restricted Group-Reflection Principle:

For any i, if $P_i(\exists j \in G(i) \subseteq R(i)\text{: } P'_j(X) \leq k) > 0$, then $P_i(X \mid \exists j \in G(i) \subseteq R(i)\text{: } P'_j(X) \leq k) \leq k$.

Here, as before, P_i stands for i's probability at a time t and P'_j stands for the probability of an agent j at some time t' that is not earlier than t. Let $R(i)$ be the set of all individuals that i considers to have the same priors as she has and to be epistemically rational. $G(i)$ is the subset of $R(i)$ consisting of all individuals j in $R(i)$ about whom i at t believes that *j at t' has all the evidence that at t is available to i and all the evidence that at t' is available to other members of R(i)* (and possibly some additional evidence as well). It is only if j satisfies these very demanding conditions that the Restricted Group-Reflection requires i to adjust her probabilities to j's probabilities.

As is easy to see, the Restricted Group-Reflection is not violated in the Story of the Hats. It's simply not applicable in that story. The agent who at stage 2 will have probability ½ for *A* will not have all the evidence that is available to other group members at that stage: she won't be seeing the color of her own hat.

It is worthwhile to reflect on the structure of our argument. In the case of the standard probability axioms, Dutch book arguments are invoked to show that a rational agent lets her degrees of belief be governed by the axioms in question. In this case we start off with seemingly reasonable principles of rational belief and the Dutch book arguments are used to shore up their intuitive appeal against a skeptic. Now, in the case of Group-Reflection, we were not quite certain what a reasonable principle would be. So we started with a naïve Group-Reflection principle. Subsequently, there are two ways of proceeding which operate in parallel.

First, we can just present counter-examples to the naïve principle and refine it by trying to duck the counter-examples. This is how we can make use of Christensen's husband-and-wife story. It teaches us that Group-Reflection has appeal only for groups whose members can rely on each other, insofar as they assume that they start from the same priors and process the evidence in the same way. Now it may well be that in the absence of these conditions we can construct a Dutch book against Group-Reflection violators. (The example of the husband-and-wife is a case in point.) But this by itself does not vindicate Group-Reflection. Compare: Suppose that someone proposes the erroneous principle that our degree of belief in the tautology should equal .9. We then show that a Dutch book can be made against an agent who violates this suggested principle by setting P(Q or not-Q) at .95 (instead of .9). Clearly, this Dutch book does not vindicate the proposed principle!

Second, we present a violation of the naïve principle in the Story of the Hats and then show that no Dutch book can be made. So now we work backwards: if no Dutch Book can be made, then maybe there is no violation of what would be an *appropriate* Group-Reflection Principle in the Story of the Hats. So what could such an appropriate Group-Reflection Principle be? In the Story of the Hats, each player at *t'* has all the evidence available to the players at *t* together with some further evidence. However, there are several players at *t'* and none of them has all the evidence available to every other player at *t'*. This means that relying on any one of them rather than on her co-players would be unreasonable. And this leads us to the presentation of the Restricted Group-Reflection Principle. Compare: Suppose that we violate the suggested erroneous principle for the tautology (the one requiring us to assign to the tautology the degree of belief of .9) by setting the degree of belief for P(Q or not-Q) at 1. Subsequently we show that no Dutch book can be made on grounds of this assignment. Then this points the way to an appropriate principle for our degree of belief in the tautology—viz. a principle which prescribes the degree of belief in the tautology to be 1 rather than .9.

The logic of Dutch Book arguments is as follows. If there is a constraint of rationality, then, in a situation in which the constraint is violated, a Dutch book can be made. So how does this help us in putting our finger on what constitutes a correct principle of rationality? First, if we have a prima facie reasonable constraint and it is possible to make a Dutch book when the constraint is violated, then this provides inductive evidence for the constraint. Second, if we have a prima facie reasonable constraint and it is *not* possible to make a Dutch book when the constraint is violated, then we know that the constraint either must be rejected or needs fine-tuning—it needs to be fine-tuned in such a way that the new and improved principle is no longer violated in the situation in which the construction of a Dutch book is precluded.

David Christensen argues against diachronic principles of rationality, including the standard individual Reflection Principle, by showing that Dutch books can be made against violators of principles that are in the neighborhood of Reflection but that are entirely implausible. One such principle is his Solidarity, which is the (unrestricted) Group-Reflection Principle applied to the case of the husband-and-wife. In this respect, diachronic principles of rationality are radically different from the probability axioms, he argues.

This is a misunderstanding of the force of Dutch book arguments.[11] The fact that a Dutch book can be made against a violator of some principle does not show that the principle is a bona fide constraint of rationality. It shows this neither for synchronic nor for diachronic principles. A Dutch book could be made against a violator of the principle that our degree of belief in the tautology should equal .9 – but this does not show that Dutch book arguments for synchronic principles of rationality should be thrown out of the window. Neither does a Dutch book against violators of the implausible Solidarity principle show that Dutch book arguments for diachronic principles of rationality should be thrown out of the window. One needs to properly appreciate the logic of these arguments: the *presence* of a Dutch book in the case of a violation of some prima facie reasonable principle suggests that this principle might turn out to be a bona fide constraint of rationality, but it does not prove it, while the *absence* of a Dutch book in the case of a violation of a prima facie reasonable principle proves that this principle is not acceptable as it stands.

So we used the absence of a Dutch book in the Story of the Hats to refine the Group-Reflection principle so that the original Group-Reflection was no longer a constraint. But where does this leave the *presence* of a Dutch book in Christensen's husband-and-wife example? The presence of a Dutch book shows that there is something epistemically amiss in the situation. But what, more precisely? One possibility is that there is a violation of Restricted Group-Reflection. This would be the case if the husband h considers his wife to belong to $G(h) \subset R(h)$ but still does not bother adjusting his degree of belief to hers. But, as Christensen tells the story, the violation of the Dutch book is due to a different kind of epistemic failing. The husband simply does not consider his wife as a member of $R(h)$: There is a lack of epistemic trust – he considers her to be unduly optimistic or pessimistic. The implication is that, in his view, she operates with different priors than he does or is not

[11] For a discussion of the difference between synchronic and diachronic Dutch books, see Rabinowicz (2008). That paper, however, does not discuss Dutch books against groups.

epistemically rational. And so the Restricted Group-Reflection does not apply. Thus, to sum up, depending on how we tell the story, a Dutch book can be made against the husband and wife either because Restricted Group-Reflection is violated or because of the lack of epistemic trust within the group.

6. What's wrong with the betting interpretation[12]

On the betting interpretation, subjective probabilities are identical with betting rates. As the Story of the Hats has shown, this interpretation is unsatisfactory. But what precisely is wrong with it?

In 'Truth and Probability', where the betting interpretation is proposed, Frank Ramsey himself mentions the limitations of this method:

> … the proposal of a bet may alter [the agent's] state of opinion; just as we could not always measure electricity intensity by actually introducing a charge and seeing what force it was subject to, because the introduction of the charge would change the distribution to be measured. (1926: 170)

There are two standard cases in which this is so. First, if you offer me a bet, then this might make me think that you have special expertise in the matter and hence I become less confident in my own judgment. In this case my betting rate matches my credence conditional on having been offered a bet. Call this the *Expert* case. Second, I may think it quite unlikely that I will, say, be able to quit smoking, but things change when I am offered a bet on that proposition. Taking a bet with high stakes on my quitting helps me to strengthen my resolve and so I become more confident that I will be able to quit. In this case my betting rate matches my credence conditional on my accepting the bet. Call this the *Smoking* case. In both cases, there is a probabilistic dependence between the proposition betted on and the availability of the bet and this explains why betting rates do not match credences.

The *Story of the Hats* is both similar and dissimilar to *Expert* and *Smoking*. In *Expert* and *Smoking*, the explanation of the disparity rests on the following claim:

(*Ramsey*) If there is probabilistic dependence between the proposition betted on and the availability of the bet, then betting rates and credences diverge.

---

[12] Our discussion in this section is influenced by Eriksson & Rabinowicz (forthcoming).

Now in *Expert* and *Smoking*, we do a simple *modus ponens* on (*Ramsey*) to explain why betting rates and credences diverge.

But this won't work in the *Story of the Hats*. On the assumption that the players settle on the symmetric trembling-hand perfect Nash-equilibrium <(0,0), (0,0), (0,0) >, nobody will ever step forward to take up the bookie's offer. So if Alice sees two hats of the same color, she is able to take up the bookie's offer, whether *A* is true or not. She knows that the bet is available. Hence, there is no probabilistic dependence between the proposition betted on (i.e. *A*) and the availability of the bet. *Modus ponens* on Ramsey fails as an explanation of why betting rates and credences diverge in the *Story of the Hats*.

The argument is slightly different. We first need to establish that, in the *Story of the Hats*,

(*Hats*) If betting rates were to match credences, then there would be probabilistic dependence between the proposition betted on and the availability of the bet.

This is easy to see. Suppose that betting rates were to match credences. Let Alice see two hats of the same color. Suppose that the proposition betted on (i.e. *A*) is true. Then no one else will step forward to sell the bet and so the chance that the bet is available to Alice is 1. Suppose that *A* is false. Then both Barbara and Carol see two hats of the same color and, on the supposition that betting rates match credences, they will also step forward, just like Alice. In that case the chance that the bet is available to Alice is 1/3. Hence, there is a probabilistic dependence between the proposition betted on and the availability of the bet.

Now we can see the logic of the argument. The argument is a *reductio.* Assume for *reductio* that, in the *Story of the Hats*, betting rates match credences. Then by (*Hats*), there is probabilistic dependence between the proposition betted on and the availability of the bet. Hence by (*Ramsey*), betting rates do not match credences. So we can reject the *reductio* premise: In the *Story of the Hats*, betting rates do not match credences.

The explanation of the divergence between betting rates and credences in the *Story of the Hats* is (*Ramsey*). In that respect it is similar to *Expert* and *Smoking*. But the argument is a *reductio* argument and not a simple *modus ponens*. In that respect it is dissimilar.

7. Conclusion

In the Story of the Hats, there is a group of epistemically ideal players with common priors, common goals, mutual trust in each other's rationality and common knowledge thereof. We construct a simple principle of Group-Reflection and show how van Fraassen's Dutch book construction against violators of the Reflection principle can be adapted to the group violating the Group-Reflection principle in the Story of the Hats.

However, if we pay attention to strategic considerations and approach the problem as a game-theorist would, it turns out that the Dutch book in the Story of the Hats is spurious. Nonetheless there is a violation of the Group-Reflection principle despite the fact that the players in our story are epistemically ideal. This induces us to restrict the Group-Reflection principle so that there is no longer a violation of the Restricted Group-Reflection in the Story of the Hats.

We take a stand against Christensen who presents a simple case of a failure of Group-Reflection and argues that Dutch book arguments are suspect for principles of diachronic rationality. We argue that Dutch book arguments do signal an epistemic deficit and show how the presence and absence of Dutch Books is instrumental in refining principles of rationality.

The Story of the Hats poses a challenge to the fair betting-rate interpretation of subjective probability. There are already well-known challenges to this interpretation, e.g. cases in which there is probabilistic dependence between the availability of a bet and the proposition on which the bet is to be made. In such cases, fair betting rates match not subjective probability *tout court*, but rather the subjective probability that the proposition would be true conditional on the availability of the bet.[13] The logic of the challenge in the Story of the Hats is different, though: the probabilistic dependence between the availability of a bet and the proposition on which the bet is to be made enters in only within the context of a *reductio* argument on the assumption that (unconditional) subjective probability determines the fair betting rate.

The Story of the Hats is a story of paradox gained and paradox lost. However, there is philosophical enchantment in this gain and loss. We have come to understand better the role of strategic considerations in betting against a group, the possibility of a reasonable principle of Group-Reflection, the place of Dutch book arguments with respect to epistemic ideals, and the betting interpretation of subjective probability.[14]

---

[13] This is still somewhat oversimplified. In cases such as Smoking, in which the bet might causally influence the proposition betted on, the rate of a fair bet instead coincides with the subjective probability that the proposition would be true if the bet were made, conditional on the availability of the bet. See Eriksson & Rabinowicz (forthcoming)

[14] We are indebted to Chlump Chatkupt for his help with preparing the final version of this paper.

References

Bovens, L. (1995), '"P and I Will Believe that Not-P": Diachronic constraints on rational Belief,' *Mind*, 104, 737-60

Bovens, L. and Rabinowicz, W. (2009) 'A Dutch Book for Group-Decision Making,' in Benedikt Löwe, Eric Pacuit, Jan-Willem Romeijn (eds.), *Foundations of the Formal Sciences VI: Reasoning about Probabilities and Probabilistic Reasoning*, Studies in Logic, vol. 16, London: College Publications UK, 91 - 101.

Bovens, L. and Rabinowicz, W. (2010) 'The Puzzle of the Hats,' *Synthese* 172, 2010, 57 - 78.

Christensen, D. (1991), 'Clever Bookies and Coherent Beliefs,' *The Philosophical Review* 100, 229-47.

Eriksson, L. & Rabinowicz, W. (forthcoming), 'The Interference Problem for the Betting Interpretation,' to appear in *Synthese*.

Lewis, D. (1980), 'A Subjectivist's Guide to Objective Chance,' reprinted with postscripts in D. Lewis, Philosophical Papers, vol. II, Oxford: Oxford University Press, pp. 83-132.

Rabinowicz, W. (2008), 'Pragmatic Arguments for Rationality Constraints,' in M.-C. Galavotti, R. Scazzieri and P. Suppes (eds), *Reasoning, Rationality and Probability*, Stanford: CSLI Publications/The University of Chicago Press, 139-63.

Ramsey, F. (1926) 'Truth and Probability,' in Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays*, Ch. VII, pp. 156–98, ed. By R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Campany. 1999 electronic edition. http://socserv.mcmaster.ca/econ/ugcm/3ll3/ramseyfp/ramsess.pdf (accessed April 20, 2011)

Robinson, S. (2001), 'Why Mathematicians Now Care about their Hat Color,' *New York Times–Science*, April 10.

Selten, R. (1975), 'A reexamination of the perfectness concept for equilibrium points in extensive games,' *International Journal of Game Theory* 4, 25-55.

Van Fraassen, B. (1984), 'Belief and the Will,' *Journal of Philosophy* 81, 235-256.