The Puzzle of the Hats

Rabinowicz, Wlodek; Bovens, Luc

Published in:
Synthese

DOI:
10.1007/s11229-009-9476-1

2010

Synthese
DOI 10.1007/s11229-009-9476-1

The puzzle of the hats

Luc Bovens · Wlodek Rabinowicz

Received: 13 February 2008 / Accepted: 31 August 2008

Abstract The *Puzzle of the Hats* is a betting arrangement which seems to show that a Dutch book can be made against a group of rational players with common priors who act in the common interest and have full trust in the other players' rationality. But we show that appearances are misleading—no such Dutch book can be made. There are four morals. First, what can be learned from the puzzle is that there is a class of situations in which credences and betting rates diverge. Second, there is an analogy between ways of dealing with situations of this kind and different policies for sequential choice. Third, there is an analogy with strategic voting, showing that the common interest is not always served by expressing how things seem to you in social decision-making. And fourth, our analysis of the *Puzzle of the Hats* casts light on a recent controversy about the Dutch book argument for the Sleeping Beauty.

Keywords Puzzle of the Hats · Credences · Betting rates · Sequential choice · Strategic voting · Dutch Book · Sleeping Beauty · Rational choice · Rationality

Our Puzzle of the Hats is inspired by a mathematical puzzle introduced in 1998 by Todd Ebert. See Robinson (2001).

L. Bovens (✉)
Department of Philosophy, Logic and Scientific Method, London School of Economics,
Houghton Street, London WC2A 2AE, UK
e-mail: L.Bovens@LSE.ac.uk

W. Rabinowicz
Department of Philosophy, Lund University, Kungshuset, Lundagård,
222 22 Lund, Sweden
e-mail: Wlodek.Rabinowicz@fil.lu.se

Published online: 28 February 2009

1 The *puzzle of the hats* and a Dutch book for a group of rational players

We randomly distribute white and black hats to a group of three rational players. We do this in the dark, with each player having an independent fifty–fifty chance of receiving a hat of one colour or the other. Clearly, the chance that

(A) not all hats are of the same colour

is 3/4: Out of eight possible hat distributions among three players, there are six in which it is not the case that all the players get hats of the same colour. Subsequently, when the lights are turned on, each player can see the colour of the hats of the other players but not of his own hat. Then no matter what combination of hats was assigned, at least one player will see two hats of the same colour. If (A) is false, this will apply to every player, and if (A) is true, there will be one player in this epistemic situation. For her, the chance that not all hats are of the same colour strictly depends on the colour of her own hat and hence equals 1/2. (Remember that the colour of each hat has been decided by an independent lottery.)

On Lewis's principal principle, a rational player will let her credences be determined by these chances. So initially all players will assign credence of 3/4 to (A), and after the light is turned on at least one player will assign credence of 1/2 to (A). Suppose that, before the light is turned on, a bookie offers to sell a single bet on (A) with stakes \$4 at a price of \$3 and subsequently, after the light is turned on, offers to buy a single bet on (A) with stakes \$4 at a lower price of \$2. Suppose, finally, that all of the above from the outset is common knowledge among the players.

If, following Ramsey, the credence equals the betting rate (i.e. the price-stake ratio) at which the player is both willing to buy and willing to sell a bet on a given proposition, then any of the players should be willing to buy the first bet and at least one player should be willing to sell the second bet. (If there are several volunteers for a given bet, we can assume that the bookie will pick out one of them at random.) Since the bets are on the same proposition, their stakes are equal and the price of the second bet is lower, the bookie can make a Dutch book—whether all hats are of the same colour or not, she has a guaranteed profit of \$1 (Table 1).

So, seemingly, the bookie has succeeded in making a Dutch book against a group of rational players with mutual trust in each other's rationality. But the fact that a Dutch book can be made against a group is a mark of some form of collective *ir*rationality. How worrisome is this? There are two possibilities. Either each player is trying to

Table 1 A Dutch book in the *Hats Puzzle*

	(A) is true: Not all hats are of same colour	(A) is false: All hats are of same colour
Bet 1	Player buys bet for \$3	Player buys bet for \$3
	Bookie pays out \$4	Bet is lost
Bet 2	Bookie buys bet for \$2	Bookie buys bet for \$2
	Player pays out \$4	Bet is lost
Payoffs	Bookie gains \$1	Bookie gains \$1

increase her own payoff. Then the Dutch book would not be too worrisome. After all, prisoner's dilemmas have a similar structure—when each player acts to her own advantage, the group payoff is suboptimal. In this case individual rationality would just not be in line with collective advantage. Or, alternatively, the players are supposed to act in the interest of the group as a whole, i.e. to maximize the group's total payoff, rather than their own winnings. In this case the Dutch book would be worrisome—it would be indicative of a breakdown in group rationality.[1]

We will focus here on the case in which the players act in the interest of the group. In order not to keep the reader in suspense, let us say straightaway that the rational course of action is not to sell the bookie the second bet and hence that no Dutch book can be made. Nonetheless, our solution to this puzzle will prove instructive in several respects.

2 The Dutch book disarmed

Let us focus on the second bet—the one that the bookie offers to buy.[2] The players are trying to maximise the payoff for the group. We need to determine the probability p_i with which a player i should step forward and offer to sell the bet if she sees two hats of the same colour. (Obviously, she shouldn't step forward to sell the bet if she sees two hats of different colours. For in that case she knows that the bet would be won by the bookie.) We assume that there is no pre-play communication. Hence, the players are symmetrically placed and so we are looking for a symmetrical Nash equilibrium $<p_a, p_b, p_c> = <p, p, p>$ for players Alice, Bob and Carol.

Suppose Alice sees two hats of the same colour. Given that information, we calculate the expected payoff of her stepping forward that Alice envisions for the group, $E_a[U(<1, p, p>)]$, and the corresponding payoff of her *not* stepping forward, $E_a[U(<0, p, p>)]$. (U is the group's payoff, while E_a stands for Alice's expectation of that payoff in this epistemic situation.[3] As the situation is symmetrical, the other players' expectations in the corresponding epistemic situation would of course be the same.) We calculate these expressions by first conditioning on the random variable S with values $S = 0$ when the hats are of different colours (i.e. when (A) is true) and $S = 1$ when all the hats are of the same colour (i.e. when (A) is false). Obviously, $P(S = 1) = 1 - P(S = 0)$. We start with:

$$E_a[U(<0, p, p>)] = E_a[U(< 0, p, p >)|S = 0]P(S = 0) + E_a[U(< 0, p, p >)|S = 1](1 - P(S = 0)) \quad (1)$$

[1] We discuss both possibilities in Bovens and Rabinowicz (2008).

[2] Thanks are due to Anthony Williams for first raising the suspicion that a rational player may not want take up the bookie's offer for the second bet.

[3] An alternative would be to consider this game from the *ex ante* perspective, i.e. to calculate the expected group payoffs of stepping forward versus abstaining if one sees two hats of the same colour in terms of the information that is available to the players *before* the light is turned on. It can be shown that the choice of the perspective doesn't matter when it comes to Nash equilibria: In this respect, the *ex ante* and the *ex post* perspective are equivalent.

Table 2 Conditional expected payoffs in the *Hats Puzzle*

i	$E_a[U(<0, p, p>)\|S=1, N=i]$	$P(N=i\|S=1)$
0	0	$(1-p)^2$
1	2	$2p(1-p)$
2	2	p^2

Since Alice sees two hats of the same colour, $P(S=0)=1/2$. Furthermore, $E_a[U(<0, p, p>)|S=0]=0$, since, if $S=0$, Alice is the only player who sees two hats of the same colour. This means that if she does not step forward, the bookie won't sell the bet. However, if $S=1$, i.e. if all hats are of the same colour, then the two other players will also see two hats of the same colour and step forward with probability p. If at least one of them will actually step forward, the bet will be sold and the bookie will lose, since (A) is false when $S=1$. Thus, if at least one of the other players will step forward, the group will gain \$2. We condition the expected payoff on the random variable $N=i$, for i being the number of the other players who step forward. So (Table 2),

$$E_a[U(<0, p, p>)|S=1] = \Sigma_{i=0,1,2} E_a[U(<0, p, p>)|S=1, N=i]P(N=i|S=1) \quad (2)$$

Consequently,

$$E_a[U(<0, p, p>)] = (0 \times 1/2) + \left(0(1-p)^2 + 2(2p(1-p)) + 2p^2\right) 1/2 = -(p-2)p \quad (3)$$

Now, we turn to:

$$E_a[U(<1, p, p>)] = E_a[U(<1, p, p>)|S=0]P(S=0) + E_a[U(<1, p, p>)|S=1](1-P(S=0)) \quad (4)$$

Again $P(S=0)=1/2$. $E_a[U(<1, p, p>)|S=0]=-2$ and $E_a[U(<1, p, p>)|S=1]=2$. So $E_a[U(<1, p, p>)]=(1/2)(-2)+(1/2)(2)=0$ for all values of p. $<1, 1, 1>$ is not a Nash equilibrium, since unilateral deviation to $<0, 1, 1>$ increases the group's payoff from $E_a[U(<1, 1, 1>)]=0$ to $E_a[U(<0, 1, 1>)] = -(1-2)\times 1 = 1$. $<0, 0, 0>$ is a Nash equilibrium, since unilateral deviation leaves the group's payoff at $E_a[U(<0, 0, 0>)] = E_a[U(<1, 0, 0>)] = 0$. It is not a strict equilibrium: A unilateral deviation from $<0, 0, 0>$ to $<1, 0, 0>$ does not increase the payoff to the group, but it does not decrease it either. For the other two players, Bob and Carol, we get the same result, of course.

We then investigate whether there are equilibria in mixed strategies. To do so, we note that if $0<p<1$, $<p, p, p>$ is an equilibrium only if the following equation holds:

$$E_a[U(<0, p, p>)] = E_a[U(<1, p, p>)]$$
$$\text{That is, } -(p-2)p = 0 \quad (5)$$

Equation 5 has no solution under the constraint $0 < p < 1$. Hence, there exists only one symmetric equilibrium, viz. $< 0, 0, 0 >$. In other words, a rational player who sees two hats of the same colour will not step forward to sell the bookie a bet, notwithstanding the fact that the rate of the bet offered by the bookie is in line with the player's credence $P(\mathrm{A}) = 1/2$.

3 An intuitive account

So why is it that a person whose credence for some proposition is 1/2 should not be posting her betting rates accordingly? Why should she refrain from expressing her willingness to sell a bet for \$2 that pays \$4? How can one explain this in intuitive terms? Since we are looking for a symmetric solution, we need to consider what would happen if every player would be willing to sell the bet under these conditions with probability $p > 0$.

We will first show that $< 1, 1, 1 >$ cannot be a rational solution. Suppose that we were all resolved to step forward and offer to accept the second bet if we see two hats of the same colour. Then I would reason as follows when seeing two hats of the same colour. There are two states—one that is favourable and one that is unfavourable for selling the bet. In the favourable state, all the hats are of the same colour. The bookie loses the bet and the group gains \$2. In the unfavourable state, the hats are of different colours. The bookie wins the bet and the group loses \$2. In the favourable state, with all hats of the same colour, two other players will step forward and nothing is lost by my not stepping forward. In the unfavourable state, with the hats being of different colours, I am the only one who would step forward, so I can save the group from a loss by not stepping forward. Hence unilateral deviation from $< 1, 1, 1 >$ to the pure strategy of not stepping forward improves the group payoff, i.e. $< 1, 1, 1 >$ is not a Nash equilibrium. We can conclude that simply stepping forward—i.e. stepping forward with probability 1—when seeing two hats of the same colour cannot be a rational solution.

Subsequently we need to understand why $< p, p, p >$ is not a Nash equilibrium for $1 > p > 0$. It is instructive to construct a graph. We have plotted $E_a[U(< 0, p, p >)]$, $E_a[U(< 1, p, p >)]$ and $E_a[U(< p, p, p >)]$ in Fig. 1. Note that for any value of $p > 0$, unilateral deviation to $< 0, p, p >$ increases the payoff function. If the others are willing to step forward when seeing two hats of the same colour or if they at least are inclined to do so with positive chance p, I can exploit this by refraining from stepping forward myself. This would decrease the chance of a win for the group in the favourable situation when all hats are of the same colour with probability $(1 - (1 - p)^3) - (1 - (1 - p)^2) = (p - 1)^2 p$. However, it would at the same time decrease the chance of a loss for the group in the unfavourable situation when the hats are of different colours with probability p. Since the win and the loss are equal in size, the probability of the occurrence of the favourable state is the same as that of the unfavourable state in the eyes of a player who sees two hats of the same colour, and $p > (p - 1)^2 p$ for $1 > p > 0$, the expected payoff to the group increases if I unilaterally deviate from $< p, p, p >$ by resolving to stay put. Hence, $< p, p, p >$ is not a Nash equilibrium for $1 > p > 0$.

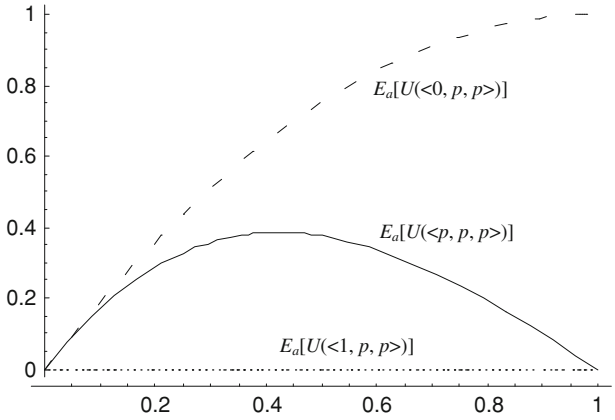

Fig. 1 Expected payoffs in the *Hats Puzzle*

Why then is $< 0, 0, 0 >$ a Nash equilibrium? This depends on two facts: (i) If it's given that none of the other players will offer to buy the bet if she sees two hats of the same colour, then my stepping forward in such situation will guarantee that I get the bet; (ii) The expected value of the bet if I see two hats of the same colour is zero, in terms of my probabilities. (i) and (ii) together imply that the expected value for the group of my unilateral deviation from $< 0, 0, 0 >$ is nil. Hence $< 0, 0, 0 >$ is a Nash equilibrium.

4 Credences as betting rates[4]

In 'Truth and Probability' Ramsey writes that 'the old-established way of measuring a person's belief is to propose a bet, and to see what are the lowest odds he will accept' (1926, p. 170). This idea has been extended in various ways. One can strengthen the relation of measurement and *identify* betting rates with credences. Or one can weaken the relation and say that it is *at least permissible* for a rational player to accept bets at betting rates that match her credences. (Bradley and Leitgeb 2006, p. 120) Furthermore, Ramsey is talking about the willingness to *buy* bets at certain rates. But one can extend this to a willingness to buy *or* sell bets at certain rates.

Ramsey is also aware of the limitations of this method: '… the proposal of a bet may alter his 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, p. 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 by taking on a

[4] Our discussion in this section is much influenced by Eriksson and Rabinowicz, mimeo.

bet with high stakes I 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 having accepted the bet. Call this the *Smoking* case. In both cases, then, there is a probabilistic dependence between the proposition betted on and the availability of the bet itself. This probabilistic dependence explains the disparity between my unconditional credence for that proposition and my betting rate.

Our case is both similar to *Expert* and *Smoking* and dissimilar from them. Let us explain. If the proposition 'Not all hats are of the same colour' is true, then Alice, who sees two hats of the same colour, will be able to sell the bet to the bookie for sure, if she steps forward to accept his offer. If, on the other hand, the proposition is false, then the chance that Alice will be able to sell the bet to the bookie is only 1/3, if she and Bob and Carol, who are in the same epistemic position as Alice, step forward to accept his offer. So, let us assume, for reductio, that the players' betting rates equal their credences for the proposition. On that assumption, there is a correlation between the truth of the proposition betted on and the probability of actually selling the bet to the bookie if one steps forward. But then, if Alice is rational, her betting rate will *not* match her credence, but rather, contrary to what has been assumed, it will match her credence conditional on the availability of the bet, i.e. conditional on the hypothesis that she would actually close the bet if she stepped forward. The latter credence is much higher than the former: By Bayes' theorem, it can be shown to be equal to 3/4,[5] which means that it would be irrational for Alice to sell the bet at the offered odds.[6] This shows that the assumption we have made cannot be upheld: In the *Hats Puzzle*, the rational players' betting rates cannot be equal to their probabilities for the proposition on which the bet is being offered. For, as we have seen, if they are, then they should not be.

However, unlike in the standard cases, in the *Hats Puzzle* the probabilistic dependence between the proposition on which the bet is offered and the availability of the bet only obtains provisionally: It only obtains under the assumption that we have made for the reductio, but not otherwise. If the players in the *Hats Puzzle* reason strategically, i.e. identify the equilibrium solution, they must come to a belief that the other players

[5] Let 'Bet' stand for the proposition that the bet is available to Alice, i.e, that she would close the bet if she stepped forward. Then

$$\mathrm{P}(\mathrm{A} \mid \mathrm{Bet}) = \frac{P(\mathrm{Bet} \mid \mathrm{A})P(\mathrm{A})}{P(\mathrm{Bet} \mid \mathrm{A})P(\mathrm{A}) + P(\mathrm{Bet} \mid \neg\mathrm{A})P(\neg\mathrm{A})} = \frac{1 \times 1/2}{(1 \times 1/2) + (1/3 \times 1/2)} = 3/4.$$

[6] Consider the following analogy. Suppose that my credence in the proposition that it will snow tomorrow at noon is 1/2. You are going to the bookmaker in town tomorrow morning and are willing to buy a bet for me on that proposition. The bet would cost me $2 and pay $4 if I win. I might refrain from letting you do so on following grounds. In the favourable state, viz. when it snows at noon, this is likely to be preceded by a cold night and there is only a small chance that your car will start in the morning. Thus you will probably be unable to drive to town and place the bet for me. In the unfavourable state, viz. when it does not snow at noon tomorrow, there is a very good chance that you will make it into town. Then clearly I will not give you the assignment to place the bet on my behalf at betting rates matching my credence in snow. My betting rate would instead match my credence that it will snow conditional on you actually placing the bet for me if I ask you to, i.e. on you actually making it to the betting office. That conditional credence is much lower than 1/2, which means that it would be irrational for me to buy the bet at the odds that are being offered.

will *not* step forward to sell the bet if they see two hats of the same colour. Given this belief about the other players, the probabilistic dependence will no longer obtain: Each player is certain that no other player will step forward, which means that the bet must be available to her—she would get it for sure if she stepped forward. Consequently, each player's unconditional probability for the proposition to be betted on is the same as her probability for that proposition conditioned on the availability of the bet.

Still, the general moral is this. If the truth of the proposition betted on is probabilistically correlated with the availability of the bet, either as things stand or at least hypothetically, i.e. under the reductio assumption that betting rates match credences, then betting rates will *not* match credences. This can be seen as a variant of a situation in which 'the proposal of a bet may alter [the agent's] state of opinion' (cf. the quote from Ramsey, above). In *Expert*, I infer from the bookie's willingness to bet that she has special expertise and hence I update my belief. In *Smoking*, I take it that the availability of a bet would give me a reason to bind myself and hence I update my belief accordingly. In the *Hats Puzzle*, I don't learn anything from the availability of the bet. But, under the reductio assumption that the players' betting rates match their credences, I would learn something from the bookie's acceptance of my offer to sell the bet he has offered to buy. And it is through a reductio argument that we come to establish that betting rates cannot match credences in the *Hats Puzzle*.

So what are the similarities and dissimilarities between the *Hats Puzzle* on the one hand and *Expert* and *Smoking* on the other hand? In all three cases betting rates diverge from credences on grounds of Ramsey's insight that (i) if there is probabilistic dependence between the proposition betted on and the availability of the bet, then betting rates do not match credences. In *Expert* and in *Smoking* we simply do a *modus ponens* on this conditional to establish that betting rates do not match credences. In the *Hats Puzzle*, the matter is more complicated, because for rational players engaging in strategic reasoning, there is no probabilistic dependence between the proposition betted on and the availability of the bet. Rather, we first establish that (ii) in the *Hats Puzzle*, if betting rates were to match credences, then there would be probabilistic dependence between the proposition betted on and the availability of the bet. Now we assume, for reductio purposes, that (iii) in the *Hats Puzzle*, the betting rates *do* match credences. We can then derive a contradiction from (i), (ii) and (iii) which permits us to reject (iii). So the core *reason* why betting rates do not match credences in the *Hats Puzzle* is the same as in *Expert* and in *Smoking*, viz. Ramsey's insight, but the *argument* to establish this is different.

Or here is another way to make our point. The complete set of reasons why betting rates and credences diverge in *Expert* and *Smoking*, consists of two facts: (a) in *Expert* and *Smoking*, there is probabilistic dependence between the proposition betted on and the availability of the bet, and (b) if there is probabilistic dependence between the proposition betted on and the availability of the bet, then there is divergence between betting rates and credences. The complete set of reasons in the *Hats Puzzle* also consists of two facts: (a′) in the *Hats Puzzle*, if betting rates and credences match, then there is probabilistic dependence between the proposition betted on and the availability of the bet, and (b) if there is probabilistic dependence between the proposition betted on and the availability of the bet, then there is divergence between betting rates

and credences. So *Expert* and *Smoking* share a reason with the *Hats Puzzle* as to why betting rates and credences diverge, but they do not share the complete set of reasons.

5 Sweetening the pie and sequential choice

Now suppose that the bookie would decide to sweeten the pie in stage 2: He proposes a rate of 2/3 for the bet he offers to buy, rather than 1/2, as in his initial proposal. Suppose that, when the lights are turned on, the bookie offers to buy a bet on the proposition 'Not all hats are of the same colour' at the rate 2/3, viz. a bet that pays \$4 and cost 2/3 of that amount, i.e. \$8/3. If it were rational for a player who sees two hats of the same colour to step forward, then a Dutch book would still be made as we can read off from Table 3 below.

But would it be rational to step forward? We will determine rational strategy profiles in this game. The calculations below can be mapped onto Fig. 2.

First, a simple 'no!' is no longer the rational solution to this game. To see this, let us look at the strategy triple $< 0, 0, 0 >$, in which each agent is resolved not to step forward. The expected payoff is clearly $E_i[\mathrm{U}(< 0, 0, 0 >)] = 0$, from the point of view of each player i. Suppose that Alice sees two hats of the same colour and she

Table 3 The *Hats Puzzle* with a Dutch book on sweetened pie betting rates

	(A) is true: Not all hats are of same colour	(A) is false: All hats are of same colour
Bet 1	Player buys bet for \$3	Player buys bet for \$3
	Bookie pays out \$4	Bet is lost
Bet 2	Bookie buys bet for \$8/3	Bookie buys bet for \$8/3
	Player pays out \$4	Bet is lost
Payoffs	Bookie gains \$1/3	Bookie gains \$1/3

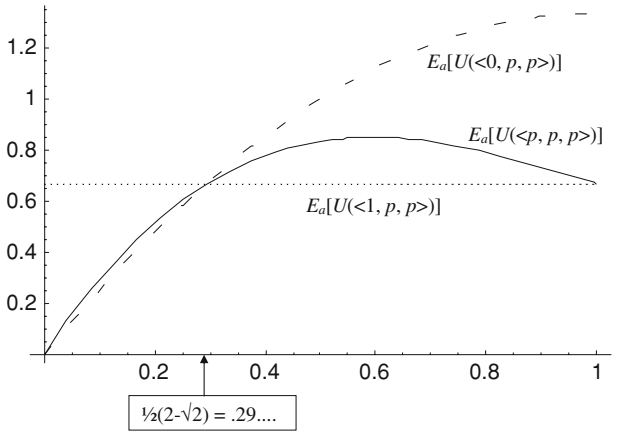

Fig. 2 Expected payoffs in the sweetened-pie *Hats Puzzle*

entertains to unilaterally deviate from this strategy. Then her expected payoff for the group is

$$E_a[U(<1,0,0>)] = E_a[U(<1,0,0>)|S=0]P(S=0) \\ +E_a[U(<1,0,0>)|S=1](1-P(S=0)) \qquad (6)$$

As before, $P(S=0) = 1/2$ from Alice's perspective. If the hats are not of the same colour (i.e. $S=0$), then the bet is won by the bookie and hence $E_a[U(<1,0,0>)|S=0] = 8/3 - 4 = -4/3$. On the other hand, if the hats are of the same colour, then the bet is lost by the bookie and hence $E_a[U(<1,0,0>)|S=1] = \$8/3$. So $E_a[U(<1,0,0>)] = (1/2)(-4/3) + (1/2)(8/3) = 2/3 > 0$, which implies that $<0,0,0>$ is not a Nash equilibrium.

But neither is $<1,1,1>$ a Nash equilibrium. Suppose everyone is resolved to step forward to accept the bet. Then, from Alice's perspective, the expected payoff for the group is $(1/2)(-4/3)+(1/2)(8/3) = 2/3$. But note that Alice can improve the payoff by unilaterally deviating. If the hats are not of the same colour and so the bet would be won by the bookie, then Alice would be the only one to step forward on the $<1,1,1>$ strategy. It would be good to avoid this: $E_a[U(<0,1,1>)|S=0] = 0$. And if the hats are of the same colour and so the bet would be lost by the bookie, then everyone would step forward on the $<1,1,1>$ strategy. So nothing would be lost by Alice's unilateral deviation: $E_a[U(<0,1,1>)|S=1] = 8/3$. Hence $E_a[U(<0,1,1>)] = (1/2)(0) + (1/2)(8/3) = 4/3 > 2/3 = E_a[U(<1,1,1>)]$.

So what are the Nash equilibria? To answer this question, we must ask for each player how she would operate on the background knowledge of seeing two hats of the same colour whilst not knowing what the other players are seeing. We check whether a profile is a Nash equilibrium by considering whether any player would have reason to deviate from it in this epistemic situation. Consider the profile $<1,0,0>$. Alice would not have reason to deviate from her strategy upon seeing two hats, since $E_a[U(<1,0,0>)] = (1/2)(-4/3)+(1/2)(8/3) = 2/3 > 0 = E_a[U(<0,0,0>)]$. Bob would not have reason to deviate from his strategy upon seeing two hats of the same colour, since $E_b[U(<1,0,0>)] = (1/2)(0) + (1/2)(8/3) = 4/3 > 2/3 = (1/2)(-4/3) + (1/2)(8/3) = E_b[U(<1,1,0>)]$. And similarly for Carol. We can construct an analogous argument for all three profiles of pure strategies in which one person steps forward upon seeing two hats of the same colour and two people do not. Hence $<1,0,0>$, $<0,1,0>$ and $<0,0,1>$ are (strict) Nash equilibria in pure strategies.[7]

Are profiles in which two people step forward and one person does not Nash equilibria? Consider the profile $<1,1,0>$. This is not a Nash equilibrium, because both Alice and Bob can improve the expected payoff by deviating from their strategy: $E_a[U(<1,1,0>)] = E_b[U(<1,1,0>)] = (1/2)(-4/3) + (1/2)(8/3) = 2/3 < 4/3 = E_a[U(<0,1,0>)] = E_b[U(<1,0,0>)]$. Hence $<1,1,0>$, $<1,0,1>$ and $<0,1,1>$ are not Nash equilibria.

[7] In the unsweetened puzzle of the hats, these asymmetric solutions in pure strategies also are available, by the way, even though they are only weak equilibria in that case.

Table 4 Conditional expected payoffs in the sweetened pie *Hats Puzzle*

i	$E_a[U(<0, p, p>)\|S=1, N=i]$	$P(N=i\|S=1)$
0	0	$(1-p)^2$
1	8/3	$2p\,(1-p)$
2	8/3	p^2

So the only equilibria in pure strategies are the asymmetric profiles in which exactly one person steps forward. But asymmetric equilibria are not accessible without pre-play communication.

Let us consider therefore whether there exists a symmetric equilibrium $<p, p, p>$. Since it can't be $<1, 1, 1>$ or $<0, 0, 0>$, it must an equilibrium in randomised strategies. This time it is enough to consider just one player, Alice, since the strategy profile is symmetric. If p is a mixed strategy, it must hold that $E_a[U(<0, p, p>)] = E_a[U(<1, p, p>)]$. Clearly, $E_a[U(<1, p, p>)] = 2/3$. We calculate $E_a[U(<0, p, p>)]$ as before, except that we need to make some modifications—see Table 4.

Hence, by Eq. 2,

$$\begin{aligned} E_a[U(<0, p, p>)] &= (0)(1/2) + \left(0\,(1-p)^2 + 8/3\,(2p\,(1-p)) + 8/3p^2\right)(1/2) \\ &= 8/3p - 4/3p^2 \end{aligned} \tag{7}$$

Now we can calculate the Nash equilibrium:

$$\begin{aligned} &E_a[U(<0, p, p>)] = E_a[U(<1, p, p>)] \\ &8/3p - 4/3p^2 = 2/3 \\ &p = 1/2\left(2-\sqrt{2}\right) \end{aligned} \tag{8}$$

So, without pre-play communication, rational players will offer to sell a bet that pays \$4 and costs \$8/3 to the bookie with probability $p = 1/2\left(2-\sqrt{2}\right) = .29289\ldots$. Let us refer to this value as p^*.

Suppose that Alice sees two hats of the same colour. We know that the expected payoff of the pure strategy profile $<1, 0, 0>$ is $E_a[U(<1, 0, 0>)] = \$2/3$. But what is the expected payoff of the strategy profile $<p^*, p^*, p^*>$ from her perspective? If the hats are of different colours, then Alice is the only one who might step forward and if she does, the loss would be $-\$4/3$. If all hats are of the same colour, then the chance that at least one person would step forward is $1-(1-p^*)^3 = .6464\ldots$ and if someone does, then the gain for the group would be \$8/3. Hence, from Alice's perspective,

$$\begin{aligned} E_a[U(<p^*, p^*, p^*>)] &= p^*\,(-4/3)\,(1/2) + \left(1-\left(1-p^*\right)^3\right)(8/3)\,(1/2) \\ &= 2/3 \end{aligned} \tag{9}$$

Note that this is the expected payoff from the perspective of a person who sees two hats of the same colour. The *ex ante* expected utilities of these profiles are different.

Ex ante, from the epistemic perspective of the players at stage 1, before the lights are turned on, there is a 3/4 chance that not all hats will be of the same colour and a 1/4 chance that all hats are of the same colour. So if we are all resolved to step forward, then the expected payoff of the strategy profile $< 1, 1, 1 >$ is as follows:

$$\begin{aligned} E_{ex\ ante}[U\ (< 1, 1, 1 >)] &= (-4/3)\ (3/4) + (8/3)\ (1/4) \\ &= -1 + 2/3 \\ &= -1/3 \end{aligned} \tag{10}$$

But if only Alice is resolved to step forward, while the other players are resolved to stay put, then there is only a 1/3 chance that someone will step forward when not all hats are of the same colour,

$$\begin{aligned} E_{ex\ ante}\ [U\ (< 1, 0, 0 >)] &= (-4/3)\ (1/3)\ (3/4) + (8/3)\ (1/4) \\ &= -1/3 + 2/3 = 1/3 \end{aligned} \tag{11}$$

The same expected payoff holds for $E_{ex\ ante}[U(< 0, 1, 0 >)]$ and $E_{ex\ ante}$ $[U(< 0, 0, 1 >)]$, i.e. only Bob or only Carol stepping forward respectively.

On the $< p^*, p^*, p^* >$ profile, at most one person will step forward, with chance p^*, if not all hats are of the same colour, and if all hats are of the same colour the chance that at least one player will step forward is $1 - (1 - p^*)^3$. So,

$$\begin{aligned} E_{ex\ ante}[U(< p^*, p^*, p^* >)] &= (-4/3)\ p^*\ (3/4) + (8/3)\left(1 - \left(1 - p^*\right)^3\right)(1/4) \\ &= 1/3\left(\sqrt{2} - 1\right) \\ &= .1381\ldots \end{aligned} \tag{12}$$

Thus, we can construct the following Table 5:

At hand is a choice situation that involves three persons. In determining what constitutes a rational action, there are three attitudes that a player can take with respect to the other persons. First, she may simply ignore their existence. The bookie offers to buy a bet that is more than fair for the player, given the player's credence after seeing two hats of the same colour. The player ignores the strategic dimensions of this choice and decides to sell. If the players are like this, they can be Dutch-booked by the bookie. Second, a player may exercise control over the other players or the players may exercise joint control. This is the case of pre-play communication. If the players have the opportunity of pre-play communication, they can designate one person who will sell the bet when seeing two hats of the same colour. Clearly, no Dutch

Table 5 *Ex Ante* expected payoffs in sweetened-pie *Hats Puzzle*

$E_{ex\ ante}[U(< 1, 1, 1 >)]$	$-1/3$
$E_{ex\ ante}[U(< 1, 0, 0 >)]$	1/3
$E_{ex\ ante}[U(< p^*, p^*, p^* >)]$	.1381…

book is possible in this case. Third, in the absence of pre-play communication, the player considers the other players to be independent decision-makers and she takes into account strategic considerations. In this case, the players will each step forward to sell the bet with probability p^* upon seeing two hats of the same colour. Again, there is no Dutch-book to be made. The bookie cannot make a profit for sure, as there is no guarantee that the second bet will be made: The probability for the bookie selling that bet is $1 - (1 - p^*)^3$, which is lower than 1.

There is an interesting analogy to be drawn with sequential choice. Think of a simple Ulysses problem. Ulysses has two choice nodes. At a later point, he can decide whether to jump in the sea upon hearing the sirens sing, or he can at an earlier point pre-empt this later choice by binding himself to the mast, so that he won't be able to jump into the water when sailing past the sirens. Now, Ulysses might most prefer hearing the sirens sing unbound and living to tell the story. However, he knows that his future self will have a different preference: Upon hearing the sirens' song, he will most prefer to jump into the sea to join the sirens in the deep. A *naïve* (myopic) chooser aims for the best outcome in terms of his current preferences and ignores that fact that the choices of his future self might undo his plans. He therefore will choose to stay unbound and when the sirens start singing he will jump to his death—the worst outcome possible given his original preferences. A *resolute* chooser is able to coordinate the actions of his present and future self. He will commit himself to future restraint and thus will not need to physically bind himself in order to stay alive. A resolute chooser can thus secure the best outcome. Finally, the *sophisticated* chooser doesn't pre-commit himself by resolution. Instead, he predicts the future preference change and adjusts his present actions accordingly: He binds himself to the mast to secure the second best outcome.

In the *Hats Puzzle*, naïve choice corresponds to the $< 1, 1, 1 >$ profile—just like the naïve chooser ignores the choices of his future self, a naïve player ignores the choices of the other players: She willingly steps forward to accept the bet if the betting rate is attractive given her credence. The resolute choice corresponds to the $< 1, 0, 0 >$ profile (or the $< 0, 1, 0 >$ or $< 0, 0, 1 >$ profiles), since this profile is the outcome of a coordination exercise. And the sophisticated choice corresponds to the $< p^*, p^*, p^* >$ profile, since this profile recognises that the choice takes place against the backdrop of choices by other selves that are beyond one's control.[8] And just like the resolute chooser does better than the sophisticated chooser, who in turn does better than the

[8] An anonymous referee has questioned this analogy between strategic reasoning and sophisticated choice. On reflection, we are inclined to agree: The analogy is not perfect. In strategic reasoning, we predict the other player's behaviour on the understanding that she acts on the basis of a prediction about our own behaviour, which requires her to predict our predictions concerning her behaviour, etc. This circularity in reasoning is absent in the case of a sophisticated agent. In order to decide how to act, the latter must predict the behaviour of her future self, but to make this prediction she need not assume that the action of the future self will be based on guesses about how the earlier self has acted. Sophisticated choice in dynamic decision problems with perfect information makes use of backward induction, which is a non-circular form of reasoning. On the other hand, though, the analogy is closer if we consider decision problems with imperfect information, in which the later self must guess the actions of the earlier self.

naïve chooser, the *ex ante* expected payoff for the group of the profiles $< 1, 0, 0 >$ ($< 0, 1, 0 >$ or $< 0, 0, 1 >$) is greater than that of the profile $< p^*, p^*, p^* >$, which in turn is greater than the expected payoff of the profile $< 1, 1, 1 >$.

6 Strategic voting and a plea for mendacity

But, one might ask, why does the *Hats Puzzle* matter? One reason is that this puzzle exhibits a structure that is very similar to that of strategic voting in groups whose members have objectives in common but not the relevant information (Banks 1999; Feddersen and Pesendorfer 1998, 1999). To that extent the puzzle goes to the core of social epistemology. Groups are called upon to act and the utility of the actions depends on the state of the world. People get private signals about what the state of the world is like and are asked to make policy recommendations on that basis. If there is a sufficiently large bloc of people in support of some action, then the group undertakes it. 'Honesty [here interpreted as voting in accordance with one's private signal] is the best policy—when there is money in it,' wrote Mark Twain. But often there is no money in it—not even from the group's perspective. The policy of voting one's private signal is like the naïve policy in sequential choice—it reflects a lack of awareness that choice happens in a setting of multiple selves.

Here is an example. Suppose that you are part of an expedition to the Democratic Republic of Congo. There is a particular area of rain forest that is of interest and each member of the expedition is sent into this area to determine whether there are okapis present. Subsequently they have to cast an independent vote on the matter. Okapis are extremely reclusive animals but they are relatively distinctive. Thus, it's hard to spot an okapi, but if you think you spotted one, then you probably did. Substantial government funds will be released to study these okapis if and only if the majority verdict of the group is that okapis are indeed present in the area. The value of such a study—if there are indeed okapis around—will be high, but if there are no okapis in the region, then the funds will be spent in vain and there will be significant negative repercussions. (Let us assume that the utility gain from funding the study rather than not funding it when there are okapis present equals the utility loss from funding the study rather than not funding it when there are no okapis present.) Should you vote no if you did not see anything resembling an okapi? This would be madness—your vote makes a difference only if the votes of the other expedition members are equally divided. This means that many other people must vote yes for your vote to count. If many of these yes-votes are sincere, the probability of the yes vote being correct is very high. Thus, it seems it is better to vote yes even if one saw nothing. But, on the other hand, if we all reason like this, then we will all vote yes even if no one has seen anything and there is no okapi around. Just like in the sweetened *Hats Puzzle*, the rational solution is to vote yes if one saw nothing with a certain probability.

Or one could turn matters around and engage in an exercise of institutional design. Then the question is, What voting procedure encourages sincerity and thereby maximizes the chances that the verdict of the group will be correct? This may be a voting procedure that has a different threshold than majority vote. In *Okapi*, the threshold for

a group verdict should be lower, considering the differential reliability of positive and negative private signals. We can also vary the utility gains and losses and the optimal threshold for a group verdict for fixed reliability levels of positive and negative private signals. This game has the same structure as (and is relevant for the study of) jury voting about the guilt or innocence of defendants.

There is a certain dependency between signals in the *Hats Puzzle* that is absent in the Okapi example. If two players see two hats of the same colour, then the third one must do so as well. But the proportion of people who observe an okapi is not a priori restricted in this way. However, apart from this difference, the formalism to solve these puzzles is roughly similar. And the moral is similar. Truthful reporting need not be in the interest of the group. Depending on (i) whether votes are cast independently or whether there is pre-play communication, on (ii) the utilities of group actions given different states of the world, and on (iii) the thresholds on individual actions before group action materialises, different standards for making reports that are not congruent with one's observations may be warranted.

The real world is of course much more complex. The model in the *Hats Puzzle* and in Okapi is binary in all respects and the boundaries are precise: Either the group action is taken or not; private signals are either on or off; individual votes are over binary options; either there is independence of votes or votes are coordinated; thresholds for group action are precise. But in the real world, there is a range of policies to choose from, say, to fight global warming. Collective action will come to pass if a vague quota of whistle-blowing on global warming is reached. There is a continuum in the strength of private signals of the dangers of global warming, etc. Still, the general lesson remains the same. The intensity with which I should blow the whistle is not just a function of the dangers of global warming according to my evidence. I need to be aware of the strategic dimensions of whistle-blowing within an institutional context. It is imperative to restrain oneself at least somewhat from stepping forward to sell the bookie a bet, even if the evidence seems to indicate that it's a good idea. It is imperative to foster an inclination towards screaming 'Okapi!' even if we did not see a thing. Similarly, a certain tendency towards whistle-blowing on global warming may be rational even if our private signal of the seriousness of this condition is weak. Certainly, there is the danger of crying wolf and the opposite danger of the Emperor's New Clothes, but the remedy is often not simple truthfulness, but rather the right degree of truthfulness. The remedy is to foster the correct inclination to let one's public voice deviate from one's private signal to the right extent, so as to maximise social utility in a context of strategic decision-making.[9]

[9] The connection between Nash equilibria and the maximisation of social utility in strategic decision-making requires extensive discussion. As noted above, we have calculated the Nash Equilibrium from the ex post perspective—i.e. from the perspective of a person who sees two hats of the same colour—but choosing the *ex ante* perspective would get us the same result: (i) the Nash Equilibria from the *ex ante* and the *ex post* perspective are the same. Now, it can also be shown that (ii) the symmetric Nash equilibrium in randomised strategies *maximises* the expected payoff of the group from the *ex ante* perspective, as compared with other symmetric profiles (cf. Bovens and Rabinowicz 2008). In forthcoming work, Bovens, Koster and Lindner hope to show that there is a clearly defined class of games—viz. a subset of the class of doubly symmetric games—in which (ii) holds in general. The Puzzle of the Hats, viewed from the *ex ante* perspective, strategic voting games and certain tragedy-of-the-commons games are instances of this class.

7 Understanding *Sleeping Beauty*[10]

In the original *Sleeping Beauty* problem (Elga 2000), Beauty is put to sleep on Sunday night. She is awakened on Monday morning, without being told which day it is, and put to sleep again later that day. A fair coin is flipped. (It is not specified when this happens—let us assume that it is flipped on Sunday night.) If the coin comes up Tails, then Beauty will be awakened on Tuesday as well and if it comes up Heads, she won't. Beauty's memories of the Monday awakening are erased—if she is awakened on Tuesday, she will have no memory of having been awakened before. Thus, her information state will be exactly as it was on Monday morning. She knows all of the above. The query is: When Beauty is awakened on Monday morning, should her credence that the coin came up Heads be 1/2, as some philosophers have suggested, or 1/3, as others have claimed?

Hitchcock (2004) presents a Dutch book argument to the effect that Beauty's credence for Heads cannot be 1/2 on the Monday awakening. Suppose that it were. Clearly, her credence for that proposition on Sunday night, before she is put a sleep, is 1/2. At the matching betting rate, she is willing to buy on Sunday a bet from the bookie on Tails that costs \$15 and pays \$30. If her credence still were 1/2 on Monday morning, she should then be willing to buy a bet on Heads that costs \$10 and pays \$20. And since she is in exactly the same epistemic state on Tuesday morning as on Monday morning, should she be awakened twice, she should be willing to buy a bet on Heads at the same rate on Tuesday morning. This arrangement gives the bookie a sure win of \$5. (If Heads is the case, the bookie's profit is \$15 from the Sunday bet minus \$10 from the Monday bet. If Tails is the case, the bookie loses \$30–\$15 on Sunday, but gains \$10 on Monday and another \$10 on Tuesday.) On the other hand, with a credence of 1/3 on Monday morning (and on Tuesday morning if she were awakened twice), Beauty would only be willing to buy a bet on Heads that pays \$30 if the price is \$10. This would make the arrangement into a fair series of bets—Beauty wins \$5 if the coin flip comes up Heads, which *ex ante*, i.e. on Sunday, has 50% probability, and if the coin flip comes up Tails, then the bookie wins \$5. To avoid the objection that the bookie on Monday (and on Tuesday, if Tails is the case) has superior knowledge, Hitchcock stipulates that the bookie goes through the same regime of awakenings as Beauty herself.

Bradley and Leitgeb (B&L in what follows) (2006) reject this line of reasoning. Hitchcock's argument presupposes that credences and betting rates coincide. However, on B&L's view, even though Beauty's credence for Heads on Monday morning still is 1/2, her betting rate for that proposition is only 1/3 and so no Dutch book can be made against her. According to B&L, betting rates do not match credences when the agent is aware that the truth of the proposition betted on is correlated with 'the size or the existence of a bet'. To see what is meant by this, let us look at some of their examples, *Forgery* and *Hallucination*, that are intended to segue into *Sleeping Beauty*.

In *Forgery*, a coin is flipped. If it comes up Tails, you will be offered a genuine bet on whether the coin came up Heads which costs \$1 and pays \$2. If it comes up Heads,

[10] The connection between the Puzzle of the Hats and the Sleeping Beauty problem was first suggested to us by Alan Hájek.

you will be offered the same bet on the same proposition but with fake money. Your notes, and the bookie's, will be switched for forgeries. Neither the bookie nor you can distinguish between the real and the fake money. You know all of the above. In *Hallucination*, a coin is flipped. If it comes up Tails, you will be offered bets on Heads at stage 1 and at stage 2 that cost \$1 and pay \$2. If the coin comes up Heads, you will be offered one such bet on Heads at either stage 1 or stage 2 and at the other stage you will instead have a hallucination that such a bet is being offered. The hallucination is qualitatively indistinguishable from the experience of a real bet. You know all of the above.

In each case your credence in Heads is 1/2. But should you accept the bets at matching betting rates? Clearly not. According to B&L, the reason is that in each of these cases the truth of the proposition that you are asked to bet on is correlated with 'the existence or the size of the bet'. Here is an interpretation of what they might mean by this. The discrepancy between credences and betting rates rests on a probabilistic dependence between the variable *Heads* (which takes the values 1 when the coin comes up Heads and 0 when the coin comes up Tails) and some other variable, for some specification of background knowledge for the probability distribution. So what is the other variable and what is the background knowledge? Let the other variable be *RealBet*, which takes the values 1 when a genuine bet is there for us to accept if we decide to do so and 0 when this is not the case. Let the background knowledge of the probability distribution P be my having a (veridical or non-veridical) experience of being offered a bet. Then there is a probabilistic dependence between *Heads* and *RealBet* both in *Forgery* and *Hallucination*. In *Forgery*, $P(RealBet = 1) = 1/2$ whereas $P(RealBet = 1|Heads = 1) = 0$ and $P(RealBet = 1|Heads = 0) = 1$. In *Hallucination*, at each stage, $P(RealBet = 1) = 3/4$ whereas $P(RealBet = 1|Heads = 1) = 1/2$ and $P(RealBet = 1|Heads = 0) = 1$.

Now, how can this probabilistic dependence lead to a divergence between credence and betting rate? The answer seems obvious: If our decision to accept the bet would be effective with a higher probability in the state in which the proposition betted on is false than in the state in which it is true, then the bet would not be fair if its rate matched our credence for the proposition in question. The lesser probability that the decision to bet is effective in the state that is favourable for the bettor than in the unfavourable state needs to be compensated by a correspondingly better betting rate. (Parenthetically, similar divergence would be to be expected if the size of the bet's stake and price rather than the availability of the bet were probabilistically dependent in this way on the proposition betted on. For example, if both the stake and the price were twice as large in the unfavourable state than in the favourable state, this difference would need to be compensated by a better rate for the bet to be fair.)

If this is the correct way to read B&L, then their analysis of why credences and betting rates diverge in *Forgery* and *Hallucination* is precisely the same as our analysis of why credences and betting rates should diverge in the *Hats Puzzle*, on the reductio assumption that the players would step forward to sell the bet if they saw two hats of the same colour. (As we remember, this assumption, while false for players who engage in strategic reasoning, would be satisfied by naïve players, who do not take into consideration the choices of others.) In the *Hats Puzzle*, the proposition betted on is 'Not all hats are of the same colour.' Let the variable *Same* take on the

value 0 when not all hats are of the same colour and 1 when all hats are of the same colour. The bookie proposes to buy a bet, but this does not mean that the player who steps forward actually will get to sell the bet to him. So let *RealBet* be 1 if the player actually would get to sell the bet if she stepped forward and 0 if she wouldn't. The background knowledge in this case consists in the fact that the player sees two hats of the same colour, that the bookie has offered to buy a bet on $Same = 0$ and that each of the other players who sees two hats of the same colour will step forward to sell it. The player's credence and her betting rate diverge, because *Same* and *RealBet* are probabilistically dependent. In the *Hats Puzzle*, $P(RealBet = 1|Same = 0) = 1$ and $P(RealBet = 1|Same = 1) = 1/3$. Given that you see two hats of the same colour, $P(Same = 1) = 1/2$. Hence $P(RealBet = 1) = (1)(1/2) + (1/3)(1/2) = 2/3$.

Now let us turn to the Dutch book for *Sleeping Beauty*. Does this analysis transfer? Let the background knowledge be, as before, that Beauty has the experience of being offered a bet. Variables *RealBet* and *Heads* are interpreted as in *Forgery* and in *Hallucination*. However, the analysis simply does not transfer. *RealBet* and *Heads* are probabilistically independent. If Beauty has the experience of being offered a bet, then the bet actually is there for the taking, whether Heads or Tails came up: $P(RealBet = 1) = 1$, $P(RealBet = 1|Heads = 1) = 1$ and $P(RealBet = 1|Heads = 0) = 1$.

There is no divergence between credences and betting rates in *Sleeping Beauty*. Betting rates are clearly 1/3, as agreed to by everyone. But so are credences. To see this, we might use the following argument, which in this form can be found in Draper and Pust (2008, pp. 283–284), but which essentially goes back to Elga (2000). There are three assumptions. First, it seems that, if Beauty were to learn upon awakening that it is Monday, then her credence in Heads would be 1/2. After all, she knows that she would be awakened on Monday in any case, whether Heads or Tails came up. Second, if Beauty were to learn upon awakening that the coin came up Heads, then her probability of it being Monday would be 1. Third, if she were to learn upon awakening that the coin came up Tails, then her probability of it being Monday would be 1/2. So, assuming that she updates by conditionalization and letting P be her probability immediately upon awakening, when she isn't yet told what day it is or what has been the outcome of the coin toss:

(i) $P(\text{H}|\text{Mo}) = 1/2$
(ii) $P(\text{Mo}|\text{H}) = 1$
(iii) $P(\text{Mo}|\text{T}) = 1/2$

with H = Heads, T = Tails and Mo = Monday. It follows from these three assumptions that $P(\text{H}) = 1/3$.[11] So upon being awakened, and—in the Dutch book version—upon being offered a bet, Beauty's credence in Heads is 1/3. Betting rates are in line with credences.

Note that if Beauty's credence in Heads upon awakening had been 1/2, contrary to what we have argued, then—given (ii) and (iii)—(i) would have to be false. Instead of equalling 1/2, her conditional probability $P(\text{H}|\text{Mo})$ would have to equal 2/3. (This can be shown using Bayes' Theorem.) Consequently, upon being told that it is Monday,

[11] Assume (1), (2) and (3). By Bayes Theorem, $\text{P}(\text{H}|\text{Mo}) = \frac{P(Mo|H)P(H)}{P(Mo|H)P(H)+P(Mo|T)(1-P(H))}$
Solve for P(H).

Beauty would have to update her credence for Heads to 2/3. In our opinion, this implication is rather difficult to swallow.

So what went wrong in B&L's argument? Do *Forgery*, *Hallucination* and *Sleeping Beauty* have something in common that might qualify as an interpretation of B&L's specification of a sufficient condition under which betting rates and credences might come apart? Well, yes, they do. Construct a variable *RealBets* which takes as its values the number of genuine bets that an agent will be offered in the game, each time in subjectively the same circumstances. Clearly, *RealBets* is a very different variable from *RealBet*. Now calculate the conditional expectation of *RealBets*. In *Forgery*, $E[RealBets|Heads = 1] = 0$ and $E[RealBets|Heads = 0] = 1$. In *Hallucination* and in *Sleeping Beauty*, $E[RealBets|Heads = 1] = 1$ and $E[RealBets|Heads = 0] = 2$. So could this be the correct interpretation of B&L's sufficient condition under which betting rates and credences diverge? Might they mean that, if the expectation of the number of real bets conditional on the truth or falsity of the proposition betted on varies, then credences and betting rates can diverge?

Well, is it a reasonable claim to make? If I know that under the unfavourable state I will get more bet offers than under the favourable state and that each time when I get the offer my subjective circumstances will be the same, then my current decision to accept the bet is evidence that I would make the same decision in the other cases as well. Consequently, my decision to accept the bet has disagreeable evidentiary bearings: It is evidence that I will make more bets under the unfavourable state (in which I get more offers) than under the favourable state. To that extent, its evidentiary value is negative.[12] Now, one might think that this negative value requires compensation by a better rate for the bet that's being offered. This would account for the divergence between betting rate and credence.

However, this reasoning will only appeal to someone who is attracted by evidential decision theory, as was pointed out by Arntzenius (2002, 57f). Causal decision theory has taught us that decisions should be evaluated in terms of their expected causal consequences and not in terms of their evidentiary bearings. My current decision to accept a bet does not cause my decisions to accept similar bets in other subjectively similar cases. Thus, its negative evidentiary value should not be considered relevant.

So, what has gone wrong in B&L's argument? If you abide by causal decision theory, then you might say that B&L confuse

(i) the dependence of the expected number of real bets on the proposition betted on with
(ii) the probabilistic dependence of the availability of a real bet on that proposition.

[12] Recall Hitchcock's Dutch book, in which the bookie offers a bet on Head to Sleeping Beauty upon each awakening. The bet costs \$10 and pays \$20. If a decision to accept the bet is a sure sign that Sleeping Beauty will accept this bet each time she awakens, then the evidentiary value of her decision equals

$$P(Heads{=}1)(E[RealBets|Heads{=}1]\times(-10{+}20)){+}P(Heads{=}0)(E[RealBets|Heads{=}0]\times(-10)){=}$$
$$P(Heads{=}1)(1\times(-10{+}20)){+}(1-P(Heads{=}1))(2\times(-10)){=}$$
$$30P(Heads{=}1)-20.$$

Thus, the evidentiary value of the decision to accept a bet at this rate is negative as long as $P(\text{Heads}) < 2/3$.

If (ii) holds, then we may expect a divergence between betting rates and credences. (ii) holds in *Forgery* and *Hallucination*, as well as in the *Hats Puzzle* (under the reductio assumption). However, in *Sleeping Beauty* (ii) does not hold, but there, as well as in the other three cases, there is something in the neighbourhood that does hold, viz. (i). But (i) is not a ground for the divergence of betting rates and credences.

Alternatively, one might say that B&L are not confused at all: Instead, they are attracted to evidential decision theory and therefore think that in cases in which (i) holds, evidentiary bearings of decisions to bet motivate a divergence between betting rates and credences. If you subscribe to evidential decision theory, then a Dutch-book argument cannot force credences into place, since betting rates and credences do diverge when (i) holds.

But B&L are not yet home free: This is not yet an argument to the effect that our credence in Heads should be 1/2 and our betting rate for Heads should be 1/3. We still have to contend with an independent argument to the effect that our credence for Heads upon awakening is 1/3—as outlined above. This argument is not blocked by evidential decision theory as such. So then, if Beauty upon awakening finds herself with credence of 1/3 for Heads, what would the counsel be of an evidential decision theorist? Well, if Beauty's actions now provide evidence for her actions in other qualitatively indistinguishable circumstances, then she should reason as follows. My choice of this particular bet is evidence of a general betting strategy and this strategy will yield one win for every two losses. So at $P(\text{Heads}) = 1/3$, a fair bet on Heads that costs \$10 must have stakes \$50, since

$$\begin{aligned} &P(Heads = 1)(50 - 10) + P(Heads = 0)2(-10) \\ &\quad = 1/3(50 - 10) + 2/3(-20) = 0. \end{aligned} \tag{13}$$

So the evidential decision theorist would advise Beauty to buy or sell bets at betting rate 1/5.

Note, however, that the bookie could exploit such betting rates and make a Dutch book against Beauty. On Sunday, he offers to sell her a bet on Heads that costs \$30 and pays \$60. On Monday and possibly Tuesday, if she would be awakened on that day as well, he offers to buy a bet on Heads that costs \$10 and pays \$50. Then whether Heads or Tails are the case, Beauty will lose \$10 (Table 6).

So accepting the independent argument for $P(\text{Heads}) = 1/3$, the evidential decision theorist would recommend betting rates that diverge from Beauty's credences, but these rates would make her vulnerable to a Dutch book.

There are two positions that avoid the Dutch book. Either one takes it to be the case that Beauty's credence in Heads after being told that it is Monday is 1/2 and one endorses causal decision theory. On this account, Beauty's credence and betting

Table 6 A Dutch book for evidential decision theorists who upon awakening assign credence of 1/3 to heads

	Sunday	Monday	Tuesday	Net
Heads	+30	−40		−10
Tails	−30	+10	+10	−10

rate for Heads upon awakening coincide and equal 1/3. Or one takes it to be the case that Beauty's credence in Heads after being told that it is Monday is 2/3 and one endorses evidential decision theory. On this account, Beauty's credence for Heads upon awakening is 1/2 and her betting rate for that proposition is 1/3. One may call this a stalemate. But our sympathies are clearly with the former package.

8 Conclusion

We have presented a *Hats Puzzle* in which seemingly rational players who act in the interest of the group but cannot make pre-play agreements appear to be vulnerable to a Dutch book. But appearances are misleading. Game-theoretical analysis shows that rational players will not accept the bookie's offer and that no Dutch book can be made. This suggests an interesting class of cases in which credences and betting rates come apart, due to a probabilistic dependence between the proposition betted on and the availability of the bet to the player. This dependence may be actual or it may enter under the reductio assumption that credences and betting rates coincide. The bookie can try to 'sweeten the pie', i.e. raise the stakes of the second bet, in order to make betting worthwhile to the players while still making a sure profit. However, this doesn't work either. The rational course of action for the players is then to do their parts in a symmetric strategy profile, in which they step forward to accept the second bet with probability lower than 1. Thus, they are again invulnerable to exploitation. This permits us to draw parallels with sequential choice. The person who is determined to take up the bookie's offer is like the myopic chooser (and will do poorly), the player who uses a pre-play agreement to coordinate her actions with the actions of the other players is like the resolute chooser (and will do very well), and the player who in the absence of pre-play communication plays her part in a Nash equilibrium is somewhat like the sophisticated chooser (and will do moderately well). We also draw a parallel to strategic voting and show that the truthful expression of one's private information in collective decision-making may not be in the group's interest. Finally, we comment on a controversy concerning the *Sleeping Beauty*. Bradley and Leitgeb (2006) identify a class of cases in which credences and betting rates diverge as in our analysis of the *Hats Puzzle*. They take the *Sleeping Beauty* to be an instance of this class. We argue that their analysis either presupposes evidential decision theory or rests on a conflation of two different kinds of dependencies.

Acknowledgements We are grateful for comments from Darren Bradley, Richard Bradley, Geoffrey Brennan, Lina Eriksson, Alan Hájek, Hannes Leitgeb, Alex Voorhoeve and Anthony Williams.

References

Arntzenius, F. (2002). Reflections on sleeping beauty. *Analysis, 62.2*, 53–62.

Banks, J. S. (1999). Committee proposals and restrictive rules. *Proceedings of the National Academy of Sciences of the United States of America, 96*(14), 8295–8300.

Bovens, L., & Rabinowicz, W. (2008). A Dutch book for group-decision making. In *Proceedings of the formal sciences VI*. Studies in Logic, College Publications (forthcoming).

Bradley, D., & Leitgeb, H. (2006). When betting odds and credences come apart: More worries for Dutch book arguments. *Analysis, 66.2*, 119–127.

Draper, K., & Pust, J. (2008). Diachronic Dutch books and sleeping beauty. *Synthese, 164*(2), 281–289.

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis, 60*, 143–147.

Eriksson, L., & Rabinowicz, W. (2008). *A Fundamental Problem for The Betting Interpretation*, mimeo.

Feddersen, T., & Pesendorfer, W. (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *The American Political Science Review, 92*(1), 23–35.

Feddersen, T., & Pesendorfer, W. (1999). Elections, information aggregation, and strategic voting. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 10572–10574.

Hitchcock, C. (2004). Beauty and the bets. *Synthese, 139*, 405–420.

Ramsey, F. (1926). Truth and probability, in Ramsey (1931) In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (Chap. VII, pp. 156–198). 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 on 18 Jan 2008.

Robinson, S. (2001). *Why mathematicians now care about their hat color*. New York Times-Science, 10 April. http://query.nytimes.com/gst/fullpage.html?res=9C00EFDB1E3EF933A25757C0A9679C8B63.