Why Betting Odds and Credences Come Apart

Owen Rees

Abstract. Betting odds and credences come apart when agents are both confused about their self-location within a possible world, and consider it possible that they may experience more than one of these *de se* possibilities consecutively either in space or time. In these circumstances, agents that act on their betting odds rather than their credences are not susceptible to Dutch Book Arguments, suggesting that betting odds - and not credences - have the functional role in rational action. The results of the Sleeping Beauty paradox, the Dr. Evil decision and the Puzzle of the Hats can be explained with reference to *de se* possibilities. These results beg the question: should we found decision theory on credences or betting odds? There are problems for both accounts.

Keywords: betting odds, credences, possible worlds, *de se* possibilities, Sleeping Beauty paradox, Dr. Evil, Puzzle of the Hats, self-location

1 Introduction

Decision theory relies on taking two of three factors of rational action and predicting the third. Either we take an agent's credence functions and their subjective utilities and infer what their rational strategies are, or we could assume they are rational, observe their strategies and infer from either their credence functions or their subjective utilities to their subjective utilities or credence functions. But what if rational action is not determined by credences, but by betting odds? Recently several authors have given examples where an agent's betting odds and credences come apart [5, 7, 11]. One problem with these results is that traditionally betting odds have been used as an indicator of an agent's credences [8, 19], because they are far more observable than credences, which are essentially just mental states. Because the two come apart, we must choose one or the other to act as a foundation for decision theory. In this paper I will show why they come apart in the Sleeping Beauty paradox [9] and the Puzzle of the Hats [5], but not in Dr. Evil's decision [10], where all of these examples include agents learning or forgetting self-locating evidence.

1.1 Betting Odds and Credences

A credence is a degree of belief that an agent holds in a proposition, for example having a credence of $\frac{1}{2}$ that a fair coin will land heads. Credences are governed by the Principal Principle [15], which states that:

$$P(A|ch[a] = x) = x \ \ when \ \ P(ch[a] = x) > 0 \ , \tag{1}$$

where P denotes an agent's credence function and $ch[a]$ denotes the objective chance of a. The betting odds an agent has in a proposition a is a ratio $x : y$ that they are indifferent to placing a bet on or against a at stake z, or not taking the bet at all, such that when they bet on a, if a then they win $\frac{xz}{y}$ and if not-a then they lose z, and when they bet on not-a, if a then they lose $\frac{xz}{y}$ and if not-a then they win z. An agent's betting odds, defined in this way, are a kind of disposition to act in a certain way, whereas a credence is a mental state that might partially determine an agent's disposition (in conjunction with their subjective utilities and degree of rationality). The betting interpretation of probability states that an agent's betting odds are a function of their credence: that if their credence in a proposition is x/y then their betting odds in it can be given by $y - x : x$. This interpretation, known as Operationalism, is to be found prominently in the work of Frank Ramsey [19] and Bruno de Finetti [8].

1.2 Non-Self-Locating Examples Where They Come Apart

Eriksson & Rabinowicz [11] give three sorts of examples where betting odds and credences come apart not due to self-locating evidence but due to salient relations between the proposition being bet on and the proposition that a bet is closed on that proposition. There are logical relations, for example where an agent bets on the proposition that they will not have made a bet between the last five minutes and the next five minutes, a bet it is impossible to win. A rational agent does have defined betting odds for this sort of proposition, $x : 0$, for any x, but they depart sharply from the credence an agent may have about whether they will or will not place a bet, which could reasonably take any value between 0 and 1. Another departure occurs when the truth of a proposition bet on is inconsistent with the possibility that the bettor may collect winnings, for example, betting on propositions like "I will not win this bet", or "I will not collect my winnings from this bet", or "The world will end before the end of the year."[1] One could argue that it is a necessary condition for a bet that it be possible to collect the winnings - otherwise it is just a donation - but either way these logical relations appear to render an agent's betting odds less expressive than their credences.

[1] Not that this stops irrational agents from making bets on these propositions. The bookmaker William Hill reportedly took £119 in bets that the world would end over the Large Hadron Collider experiment. A spokesperson said that: "We are offering a set price of 1,000,000/1 about the world ending before the end of this year and have already taken half a dozen bets, admittedly none bigger than £10 - but we are happy to permit customers to decide on their own odds, should they wish."

The second example the authors give is an evidential relation between proposition bet on and the proposition that the bet is closed. This will usually occur to an agent whenever another agent offers them a bet at certain odds. The first agent might conclude that there's a reasonable probability that the second agent is rational (i.e. seeks to maximise expected utility for herself), that she desires to win money from the offered bet, and believes that the bet she has offered is favourable to her. Depending on how strong the available evidence to believe all of these propositions is, the first agent could make a principle of deference to the second agent, revising their own belief set with respect to the offer of the bet to conclude that the probability of winning the offered bet is less than they previously thought. So an agent might reasonably have a credence in A of 0.2, but not have betting odds for A of $4:1$. Recall that betting odds (as defined as a state of indifference between betting on or against a proposition at those odds) are a kind of disposition. So if an agent walks along the street with certain betting odds, they had better be prepared to accept bets on the propositions they concern if the odds offered are higher than those betting odds (if they are equal they are rationally permitted to take either side of the bet or reject the bet altogether). But by the time a rich bookie has offered you a bet it's too late to change your betting odds, even though your credence may now have been revised accordingly. The sensible option is to maintain a range of betting odds, close to your actual credence, that you are not willing to place a bet on a proposition either way in case a bet is offered to you which revises your credence to a value outside that range. But then betting odds and credences will never correlate, at least for the agents that do not consider themselves omniscient.

Thirdly, there are causal relations. That a bet has been placed on the proposition might make the proposition more or less likely to be true, for example, an agent might have a credence of 0.2 that they will pot a snooker ball, but not be willing to stake £1000 at odds of $4:1$ that they will, as betting so much money on it could affect their credence in it, (or they could thrive under the pressure and be more likely to pot it given they have bet on it). The amount of money to be staked may also have an undesirable effect on an agent's betting odds. Too low and we risk them not caring whether they win or lose, and too high and they could be very confident and still turn down a bet at evens because the bet contains too much variance. Bernoulli [3] argued that a sum gained is worth less than an equal sum lost. If so, it's impossible to put an accurate value on the proportion of one to the other. Certainly, different people have different degrees of risk aversion, but it is difficult to tell exactly where in that range lies rationality.

All of the above departures seem to be problems for traditional credence-based decision theory, but even worse problems for betting-odds-based decision theory. Credences are intuitively stronger as a foundation because they are psychologically accessible, so when the two depart the reasonable decision theorist is more likely to side with credences. In what follows I discuss departures between betting odds and credences due to revision by self-locating evidence, and where credence-based decision theory gets it badly wrong.

2 Self-Location and *De Se* Possibilities

Self-locating evidence is distinct from non-self-locating evidence. Self-locating evidence ascribes a position to the believing agent. For example, Ernst Mach reportedly boarded a train, saw a man standing at the other end of the carriage and concluded that "That man is a shabby pedagogue." [18] At first not realising that he was seeing himself in a mirror, he only learned a proposition about the world: that a man existed at a certain space and time who was a shabby pedagogue. Although he did learn something about himself, he couldn't have known it until he identified the man he had seen with himself. For the proposition to be significantly self-locating it would end, "and I am that man."[2] David Lewis [14] explains the difference between regular evidence and self-locating evidence in terms of possible world semantics. A possible world is a spatio-temporally closed maximally consistent set of propositions. But, says Lewis, a possible world is not the same as a possibility. One agent may have several possibilities within the same possible world. Suppose there is a world that contains two omniscient gods: one that lives on top of the tallest mountain and throws down manna and one that lives on top of the coldest mountain and throws thunderbolts. They could each know every proposition about the world but not know whether they are the god on the tallest or coldest mountain.[3] Being omniscient they would know that there was only one accessible possible world, but each would have two *de se* possibilities within it.

[2] According to John Perry, who I agree with, a proposition of the form "That man is a shabby pedagogue" does have some small degree of self-locating content because it entails knowledge of the spatio-temporal distance of the referent from the referrer, or at least some means of the referrer making reference to the referent by the use of the term "that". If we pressed the narrator further on his perceptions of the pedagogue we could eventually (possibly) locate and identify the narrator by working backwards from the perspective to the perceiver. Conversely, if we are told simply, "There is a man who is a shabby pedagogue in such-and-such carriage of such-and-such train at such-and-such station at such-and-such time" we can learn nothing of the identity or location of the narrator, at least not knowingly.

[3] In fact, Lewis suggests, their omniscience about their possible world might cause them to be ignorant about their self-location as there would be nothing to discern one time or place from another as the place that they are. In contrast, I seem to have a lot of detailed knowledge about the nature of one particular room at one particular time, and much less knowledge about other rooms, places and times, and this suggests to me that I am probably located in that room at that time, i.e. here, now.

There are roughly two kinds of self-locating evidence: spatial and temporal.[4] In the following examples the agents are asked to revise by regular evidence and self-locating evidence, which causes them to consider both that more than one possible world is accessible, and that there are an unequal number of *de se* possibilities per possible world. But this is not what causes departure, as betting odds and credences come apart in the Sleeping Beauty paradox, but not in a modified Dr. Evil decision made to yield an identical space of metaphysical and epistemic possibilities as that in the Sleeping Beauty paradox.

2.1 Sleeping Beauty, Dr. Evil and the Puzzle of the Hats

Sleeping Beauty is a paragon of rationality. On Sunday night she is to be put to sleep and plied with an amnesiac which will cause her to forget future awakenings as soon as she falls asleep again. A fair coin is flipped and if heads then she is awakened on Monday night only, and if tails she is awakened on Monday afternoon, returned to her dormitive state and awakened again on Tuesday afternoon. She knows all of the above, and that therefore every awakening will be subjectively indistinguishable from her point of view. She will be asked whenever she wakes up what her credence is that heads landed. It is uncontroversial that her credence in heads before the experiment is $\frac{1}{2}$. Some authors have maintained that her credence during the experiment in heads becomes $\frac{1}{3}$, while others say that it remains $\frac{1}{2}$.[5] She receives only self-locating evidence[6], so the debate has mainly been concerned with whether self-locating evidence can update credences. If not then it might update rational betting odds.

Beauty knows that there are two possible worlds she could be a member of when she awakes: the heads-world, H, and the tails-world, T, but there are three possibilities for her, heads on Monday, H_1 (equivalent to H), tails on Monday, T_1, and tails on Tuesday, T_2. In the betting version of the game, Beauty is offered a bet on heads whenever she wakes up. Christopher Hitchcock [13] showed that if Beauty assigns betting odds of $1 : 1$ on heads before the experiment she must assign betting odds of $2 : 1$ on heads during the experiment, otherwise she will be Dutch-Booked. A Dutch-Book is made against an agent whenever they do or might accept a set of bets guaranteed to lose them money, and they are thought

[4] For examples of spatial self-location problems see Two Roads to Shrangi La in [1], the Absent-Minded Driver [2] or the case of Rudolf Lingens being lost in the library [17]. For temporal self-locations see The Prisoner in [1], or The Cable Guy Paradox [12].

[5] See [6] p.17 for a list of references for halfers and thirders.

[6] Monton [16] established that she actually forgets self-locating evidence. Suppose that a 'tay' is a length of time of twenty-four hours beginning from 6PM on one day and ending on 6PM on the next day. When Beauty goes to sleep to begin with, she knows that it is the first tay. Whenever she wakes up she no longer knows whether it is the first or second tay, and so has become more ignorant about her temporal location and therefore her self-location.

to be indicators of an agent's incoherent credences. Fig. 1 shows Beauty's epistemic situation whenever she awakes during the experiment.

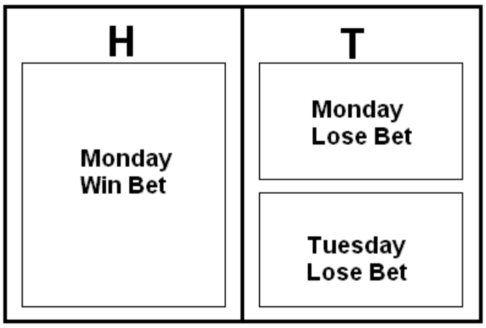

Fig. 1. Sleeping Beauty paradox. The boxes drawn with bold lines represent possible worlds, H and T. The boxes drawn with thin lines represent *de se* possibilities within those possible worlds, H_1, T_1 and T_2.

A strong argument that Beauty's credence in heads becomes $\frac{1}{3}$ was originally given by Elga [9] and has been repeated by others since. Given that Beauty will wake up on Monday regardless of whether heads or tails landed, the experimenters could decide to wake her up on Monday, let her go back to sleep, and only then flip the coin to determine whether they ought to wake her again on Tuesday or not. So if she awakes and is told that it is Monday, she ought to maintain that the probability of the coin landing heads is $\frac{1}{2}$, as it may not have even been flipped yet, i.e. $P(H|Monday) = \frac{1}{2}$. But Fig. 1 suggests that her credence in heads upon being told it is Monday ought to be $\frac{2}{3}$, given that if she were told tails had landed she would have to assign a credence of $\frac{1}{2}$ to being in either T_1 or T_2. Her credence in H would then be given by the probability of H divided by the disjunctive probability of H or T_1: $\frac{1}{2} \div (\frac{1}{2} + \frac{1}{4})$. Given that it is Monday, it is argued, we can rule out T_2 as a *de se* possibility, but actually we cannot. If T_1 is true, then T_2 *will be true*, and if T_2 is true then T_1 *has been true*. Being told that it is Monday doesn't teach Beauty anything about the outcome of the coin. The trick in the Sleeping Beauty paradox is that while the combined probabilities of H and T sum to 1, the combined probabilities of H_1, T_1 and T_2 sum to 1.5, equal to the mean number of awakenings Beauty expects to experience. It is because Beauty can be consecutively more than one of the *de se* possibilities that her betting odds and credences depart, because she can experience both T_1 and T_2 consecutively and expect to bet on both occasions. A bet made before the experiment or on Monday is always fair, but if she makes a bet on Tuesday then not only is it no longer fair (it is unwinnable), but the stake is doubled to account for the bet she made on the same proposition on Monday. To show that betting odds and credences come apart due to the possibility of multiple self-locations, and not due to a disproportionate ratio of possibilities per possible world over the range of possible worlds, we shall consider a modified Dr. Evil decision comparatively similar to the Sleeping Beauty paradox but where Dr. Evil cannot be more than one of his *de se* possibilities consecutively.

Dr. Evil mans an impregnable fortress on the Moon and is planning to imminently destroy the Earth. Supposing that we can convince him fully that we have a brain-in-a-vat on Earth experiencing exactly the same mental life as Dr. Evil, we could persuade him not to destroy the Earth and in return we will not torture the brain. Dr. Evil values his own non-torture more than he values the destruction of the Earth. If he knows that there is a brain-in-a-vat experiencing exactly what Dr. Evil is experiencing then he ought to assign equal credences to being Dr. Evil and being the brain, i.e. $\frac{1}{2}$. Suppose that for all Dr. Evil knows, there is only a $\frac{1}{2}$ chance that the brain-in-a-vat will work and has been turned on already, he will consider there to be two possible worlds - call them H and T - where the brain-in-a-vat does not work and where it does work respectively, and three *de se* possibilities for him - call them $Dr.Evil_1$, equivalent to H, and $Dr.Evil_2$ and BIV, both contained within T. Fig. 2 represents Dr. Evil's epistemic position.

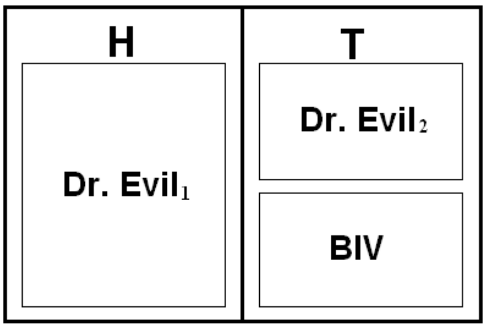

Fig. 2. Probabilistic Dr. Evil.

One difference between the Sleeping Beauty paradox and the modified Dr. Evil is that the former constitutes a temporal self-location problem and the latter constitutes a spatial self-location problem, where Dr. Evil does not know whether he is located on the Moon or in a vat in our laboratory. Only in very rare cases will a spatial self-location problem lead to a departure between betting odds and credences, because it's a lot easier to imagine examples where an agent is the same person at the same location at different times than it is to imagine examples where an agent is the same person at the same time at different locations. The second difference is that Dr. Evil cannot be any two of the possibilities jointly. He can't both set off the missiles and be the tortured brain. The combined probability of all his *de se* possibilities sums to 1, and for that reason he can work out that the probability of his being the brain-in-a-vat is $\frac{1}{4}$, and at no point do his betting odds and credences depart. Suppose that we had activated nine brains-in-vats with certainty, so that Dr. Evil's credence that he is Dr. Evil is only $\frac{1}{10}$. It would be impossible for him to experience more than one torturing, whereas the equivalent version of the Sleeping Beauty paradox would have a $\frac{1}{2}$ probability that Beauty is awakened nine times to make nine bets and cause her to demand odds of $9:1$ on any bet that she does make. Whenever the combined probability of an agent's *de se* possibilities sum to greater than 1 her betting odds and credences will come apart.

As mentioned, it is rare for a spatial self-location problem to result in a departure, but this is what happens in the Puzzle of the Hats [5], where three rational agents are jointly Dutch-Booked if they use their credences to generate their betting odds. Each agent is wearing a hat, either black or white, with each hat having an independent chance of $\frac{1}{2}$ of being either colour. They are asked to consider the proposition, X, that not all hats are of the same colour. Before the lights are turned on none of them can see the colour of any hats, so they each calculate the probability of X as $\frac{3}{4}$ $(1 - P[\text{all black}] - P[\text{all white}])$. When the lights are turned on they will each be able to see the hat colours of the other two agents. $\frac{3}{4}$ of the time only one of them will observe two hats of the same colour, and therefore believe that the truth of X depends upon the colour of her own hat and so change her credence in X to $\frac{1}{2}$, and $\frac{1}{4}$ of the time all three agents will observe two hats of the same colour and so will all change their credences in X to $\frac{1}{2}$. Bovens & Rabinowicz argue that a Dutch-Bookie can sell the trio a single bet on X before the lights are turned on for £3 returning £4 if successful, and then after the lights are turned on buy back the same bet for just £3 because there will be at least one agent who will now consider the probability of X to be $\frac{1}{2}$. The Dutch-Bookie is guaranteed to make a £1 profit if the rational agents act according to their credences. Fig. 3 represents the agents' prior probabilities for being in certain positions after the lights are turned on.

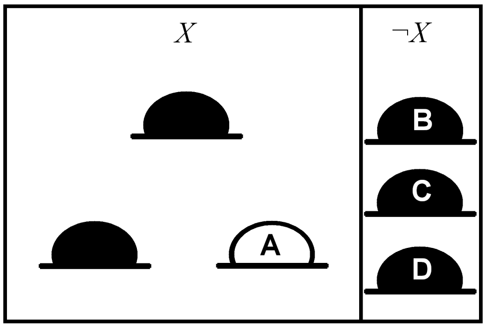

Fig. 3. The Puzzle of the Hats. Given that you are seeing two hats of the same colour there are two salient possible worlds, X and $\neg X$ and four *de se* possibilities for you, A, B, C and D.

A, B, C and D are the possibilities available to those agents who can see two hats of the same colour. Given that they can see two hats of the same colour, the probability that they are A is $\frac{1}{2}$, and $P(B) = P(C) = P(D) = \frac{1}{6}$.[7] So each agent in this position realises that half of the time his bet will win and half of the time it will lose. The Dutch-Book consists in the fact that whenever the bet is profitable (i.e. the quarter of the time when X is false), all three of the agents will step forward to make the bet but only one bet will be accepted. There's a causal relation between the truth of the proposition to be bet on and the availability of the bet, in favour of the bookie.[8] Whenever you're in a position to step forward to take the bet, half the time you do so and lose. The other half of the time, $\frac{1}{3}$ of the time you are the first forward, your bet is accepted and you win, and $\frac{2}{3}$ of the time you are beaten by one of the others and you don't get to place a bet at all. How is this a spatial self-location problem? Well, all three of the agents are rational, so they play the same strategy. This means that they all know that if they play a certain way then the others are playing the same way, so there is really only one decision being made between the three of them. It doesn't matter whether they are playing with the same bankroll or with individual bankrolls. As in the Sleeping Beauty paradox, as a group they realise that they are either going to be A and lose a bet $\frac{3}{4}$ of the time or they are going to be B, C and D *together* $\frac{1}{4}$ of the time and win one bet. So they ought to demand odds of $3:1$ on X after the lights are turned on even when their credence in X

[7] The prior probability of A is $\frac{3}{4} \times \frac{1}{3}$, and the prior probabilities of B, C and D are each $\frac{1}{4} \times \frac{1}{3}$. So $P(A|\text{I am seeing two hats of the same colour}) = \frac{1}{4} \div [\frac{1}{4} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12}] = \frac{1}{2}$, and $P(B|\text{I am seeing two hats of the same colour}) = \frac{1}{12} \div [\frac{1}{4} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12}] = \frac{1}{6}$.

[8] For another Dutch-Book that uses this method see [4].

will only ever be $\frac{1}{2}$ or 1 (when they can see two hats of different colours). The combined probabilities of their *de se* possibilities A, B, C and D sum to 1.5, and this is why betting odds and credences depart in the Puzzle of the Hats.

3 Betting Odds Versus Credences

I have shown that there are two ways that betting odds and credences depart, and that there are problems for any decision theory which relies on just one of either betting odds or credences. Making decisions based on betting odds always results in utility maximisation, but they are not psychologically available to agents. If anything they can only be calculated as a function of credences, and in the cases where the credences generate bad betting odds, we can only work out the right betting odds by working backwards from the most profitable decision to see what they should be! Most non-theoretical agents do not consider what their betting odds are for propositions unless they are making real bets on them with money, so it would be difficult to apply a decision theory based on betting odds to real people. Also, as is pointed out in [11], the whole point of using betting odds to begin with was as an observable indicator of an agent's credences. If we restrict betting-odds-based decision theory to theoretical agents, then we have little need for them as we can postulate all the credences we like. The optimal solution would be to eliminate betting odds as much as possible, and keep the distinction between possible worlds and *de se* possibilities. As possible worlds are to credences, so *de se* possibilities should be to something else - *de se* credences. It's obvious from the literature on the Sleeping Beauty paradox that many authors believe that Beauty's credence in heads becomes $\frac{1}{3}$ when really this is her *de se* credence, the probability of her experiencing a heads-world divided by the combined sum of the probabilities of all of her *de se* possibilities, which happen to add to more than 1. The distinction between them is subtle, and it is easy to conflate the two. I do not think that applying both kinds of credence is a violation of the Principal Principle, which establishes a relation between subjective and objective chances. Credences and *de se* credences are two kinds of subjective chance, which is why they are both crucial for successful decision theory.

References

1. Arntzenius, F.: Some Problems for Conditionalization and Reflection. Journal of Philosophy. 100, 356–370 (2003)
2. Aumann, R., Hart, S., Perry, M.: The Absent-Minded Driver. Games and Economic Behavior. 20, 102–116 (1997)
3. Bernoulli, D.: Exposition of a New Theory on the Measurement of Risk. (1738) Reprinted in Econometrica. 22, 23–36 (1954)

4. Bovens, L.: Four Brides for Twelve Brothers How to Dutch Book a Group of Fully Rational Players. In: Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz. Rønnow-Rasmussen, T., Petersson, B., Josefsson, J., Egonsson, D. (eds) (2007)
5. Bovens, L., Rabinowicz, W.: The Puzzle of the Hats. Synthese. 172(1), 57–78 (2010)
6. Bradley, D.: How Belief Mutation Saves Conditionalization from Self-Locating Information. `http://faculty.arts.ubc.ca/dbradley/works/RLT8.pdf` In Preparation
7. Bradley, D., Leitgeb, H.: When betting odds and credences come apart: More worries for Dutch book arguments. Analysis. 66, 119–127 (2006)
8. de Finetti, B. Foresight. Its Logical Laws, Its Subjective Sources. (1936) Reprinted in Studies in Subjective Probability. Kyburg, Jr., H., Smokler, H.(eds.): John Wiley & Sons, Inc., New York (1964), 93–158
9. Elga, A.: Self-locating belief and the Sleeping Beauty problem. Analysis. 60, 143–147 (2000)
10. Elga, A.: Defeating Dr. Evil with self-locating belief. Philosophy and Phenomenological Research. 69, 383–396 (2004)
11. Eriksson, L., Rabinowicz, W.: A fundamental problem for the betting interpretation. In Preparation
12. Hájek, A.: The Cable Guy Paradox. Analysis. 65, 112-119 (2005)
13. Hitchcock, C.: Beauty and the bets. Synthese. 139, 405–20 (2004)
14. Lewis, D.: Attitudes *de dicto* and *de se*. 133–156 (1979) Reprinted in Philosophical Papers Volume I. Lewis, D., New York, Oxford University Press (1983)
15. Lewis, D.: A Subjectivist's Guide to Objective Chance. 114-132 (1980) Reprinted in Philosophical Papers Volume II. Lewis, D., New York, Oxford University Press (1986)
16. Monton, B.: Sleeping Beauty and the Forgetful Bayesian. Analysis. 61, 47–53 (2001)
17. Perry, J.: Frege on Demonstratives. Philosophical Review. 86, 474–497 (1977)
18. Perry, J.: Myself and I. 83–103 (1997) In Philosophie in Synthetisher Absicht (A festschrift for Dieter Heinrich). Stamm, M. Klett-Cotta, Stuttgart (1998)
19. Ramsey, F.: Truth and Probability. 19–47 (1926). Reprinted in Decision, Probability and Utility. Gärdenfors, P., Sahlin, N. (eds.), Cambridge University Press (1988)