

ON THE REFLECTION PRINCIPLE

A Master's Thesis

by
GİZEM KAVAS

Department of
Philosophy
İhsan Doğramacı Bilkent University
Ankara
June 2018

GİZEM KAVAS

ON THE REFLECTION PRINCIPLE

Bilkent University 2018

To Mustafa



ON THE REFLECTION PRINCIPLE

The Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

GİZEM KAVAS

In Partial Fulfilment of the Requirements for the Degree of
MASTER OF ARTS

THE DEPARTMENT OF
PHILOSOPHY
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA

June 2018

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Philosophy.

SWU

Assoc. Prof. Dr. Simon Wigley (Bilkent University)
Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Philosophy.

W.G. Wringe

Asst. Prof. Dr. William G. Wringe (Bilkent University)
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Philosophy.

L. Thorpe

Asst. Prof. Dr. Lucas Thorpe (Boğaziçi University)
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences.

Halime

Prof. Dr. Halime Demirkan
Director

ABSTRACT

ON THE REFLECTION PRINCIPLE

Kavas, Gizem

M.A., Department of Philosophy

Supervisor: Assoc. Prof. Dr. Simon Wigley

June 2018

In ‘Belief and the Will’ van Fraassen (1984: 255) argues that traditional epistemology allows one to believe in propositions that are not entailed by the evidence at hand, while at the same time acknowledging that what is taken to be evidence may in fact turn out to be false. However van Fraassen (1984) shows the incoherence in such agent’s belief states by a diachronic Dutch Book argument, in which such agents are shown to be vulnerable to sure losses in sets of wagers which he regards as fair, which deems him as irrational. As a solution to this problem, van Fraassen (1984) formulates the Reflection Principle as a further constraint for rational agents’ credences. In this thesis I will critically evaluate the Reflection Principle as a constraint for rational agent’s credences, and conclude that the Reflection Principle as an attack and solution to the traditional epistemological view is unsuccessful since it is incompatible with epistemic agency.

Keywords: Credence, Dutch Book Arguments, Rationality Constraints, Reflection Principle, Sleeping Beauty Problem

ÖZET

YANSIMA PRENSİBİ ÜZERİNE

Kavas, Gizem

Yüksek Lisans, Felsefe Programı

Tez Yöneticisi: Doç. Dr. Simon Wigley

Haziran 2018

Van Fraassen (1984: 255), ‘Belief and the Will’ adlı eserinde geleneksel epistemolojinin, kişiye sahip olduğu bulgularca gerektirilmeyen önermelere inanması için imkan verdiğini ve aynı zamanda bulgu olarak kabul edilenlerin de yanlış olarak ortaya çıkabilme ihtimalini göz önünde bulundurmaktadır. Bununla birlikte van Fraassen (1984) kişilerin inanç sistemlerindeki tutarsızlığı artzamanlı Dutch Book argümanları ile göstermektedir. Bu argümanlarda, kişiler, makul buldukları bahisleri kaybetmekte ve irrasyonel oldukları tespit edilmektedir. Bahsedilen probleme çözüm olarak, van Fraassen (1984) Yansıma Prensibi’ni, rasyonel ve dereceli inanç sistemleri üzerine ek bir kısıtlama olarak formüle etmektedir. Bu tezde, rasyonel kişilerin dereceli inanç sistemleri üzerine getirilen Yansıma Prensibi kısıtlaması, eleştirel açıdan değerlendirmeye tabi tutulacaktır. Sonuç olarak, geleneksel epistemolojiye karşı çıkan Yansıma Prensibi epistemik faillik kavramı ile uyumsuz olduğu için yetersiz bulunacaktır.

Anahtar Kelimeler: Dereceli İnanç Sistemleri, Dutch Book Argümanları, Rasyonalite kısıtlamaları, Uyuyan Güzel Problemi, Yansıma Prensibi

ACKNOWLEDGMENTS

I would like to thank Bill Wringe for his continuous support and guidance. I feel like I am so lucky to have known a person like him who has such in depth and breadth knowledge about almost anything! I am especially grateful to him for suggesting me the epistemic agency challenge.

I would like to express my gratitude to Simon Wigley for being a patient and supporting advisor.

I would also like to thank Sandrine Berges for being a great example that she is. Her efficiency and brightness is exceptional and inspiring.

I am also very grateful to my family, for being there for me when the stakes are high and respecting my agency.

But above all, I can not thank enough to Mustafa Yıldız for merely being the person that he is. You have always been there for me with everything you have.

I love you so much, you are the best!

TABLE OF CONTENTS

ABSTRACT	iii
ÖZET	iv
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	vii
CHAPTER 1: INTRODUCTION	1
1.1 Setting	2
1.2 Dutch Book Arguments	3
1.3 Diachronic Dutch Book Argument	7
1.4 Van Fraassen’s Diachronic Dutch Strategy	7
CHAPTER 2: SLEEPING BEAUTY	11
2.1 Sleeping Beauty Problem	11
2.2 Elga’s Argument and It’s Consequences for the Reflection Principle	13
CHAPTER 3: ALTERNATIVE EXPLANATIONS	19
3.1 Self Locating Beliefs Count as Evidence	21
3.2 Lessons from Hitchcock and Elga	27
3.3 Credences and Fair Betting Odds Coming Apart	28
CHAPTER 4: REJECTING THE REFLECTION PRINCIPLE	37
REFERENCES	44

LIST OF TABLES

1	Van Fraassen's Diachronic Dutch Strategy	8
2	Computed Values for Bets	9
3	Forgery Scenario 1	29
4	Forgery Scenario 4	31
5	Hallucination Scenario	33
6	Sleeping Beauty Bets	34



CHAPTER 1

INTRODUCTION

In this thesis I will critically evaluate van Fraassen's (1984) Reflection Principle as a constraint for rational agent's credences. For that purpose, I will discuss Elga's (2000) Sleeping Beauty Problem and its consequences for the Reflection Principle. In the light of the discussions, I will argue that the Reflection Principle is problematic since it enforces complete credence functions on agents. I will conclude that the Reflection Principle's incompatibility with epistemic agency provides grounds for rejecting the Reflection Principle.

In this chapter I will firstly discuss the setting in which van Fraassen (1984) motivates the Reflection Principle. Secondly, I will discuss the way he motivates his Reflection principle with diachronic Dutch Book arguments (van Fraassen, 1984). However van Fraassen's diachronic Dutch Book arguments assume an understanding of the Dutch Book arguments. For that reason, I will first provide a clarificatory discussion of what Dutch Book arguments are. Thirdly I will discuss the Reflection principle itself (van Fraassen, 1984).

1.1 Setting

Van Fraassen in ‘Belief and the Will’ (1984) merely aims to save traditional epistemology from the undesirable results that it yields with respect to it allowing irrational belief formations. One is in the irrational belief state van Fraassen (1984: 237) attacks, when one allows himself to believe in a proposition which is not entailed by the previous propositions or evidences, while at the same time believing that she may end up believing in a proposition which may in fact be false . However van Fraassen (1984: 237- 238) claims that such belief states should be disallowed since they are incoherent and thus irrational.

In order to illustrate the incoherency, van Fraassen (1984: 237) gives an example of a person who, for the sake of this argument, is assumed to not have an opinion about the theory of evolution as of today. For the purposes of ease of explanation, let the name of this person be Amanda. So, in addition to not having any opinion about the theory of evolution today, it seems quite likely to Amanda that some time in the future like the upcoming two years, she can end up being certain of the theory of evolution. Since after all she has just enrolled for an undergraduate program in biology as of today. However, being the reasonable person that she is, Amanda also believes that it is quite possible for her to be certain of the theory of evolution, while the theory of evolution being in fact false. Perhaps Amanda has perceived herself in states where she was certain of the truth of a proposition only to later find out that she was in fact certain of a proposition which was in fact false. Van Fraassen (1984) claims that Amanda, or anyone with an equivalent set of beliefs for that matter has incoherence among her beliefs. Van Fraassen (1984) formalizes such incoherence among beliefs with Dutch Book arguments. However since van Fraassen’s

(1984) discussion assumes an understanding of Dutch Book arguments, I first discuss what Dutch Book arguments are.

1.2 Dutch Book Arguments

Ramsey (1926: 22) states that rules of probability applies to any consistent set of credences. If any (set of) belief(s) is violating these rules, then a book can be made against this belief(s). Similarly, Hájek (2008: 173) claims that Dutch Book arguments illustrate that credences should be consistent with the probabilistic reasoning. These arguments are made in such a way that agent's credences are in line with the ticket prices for bets they regard as fair (Hájek, 2008: 173). If you assign $1/2$ credence for a coin toss landing Heads, then you would regard a bet fair if the ticket price for entering that bet is 5 dollars, while the bet pays 10 dollars if Heads and nothing otherwise. So a generalization of this claim can be made in the following way:

Let H denote the outcome of the coin toss to be Heads, and $P(H)$ be your credence for H . Then, a fair bet for H that pays M when H holds, would be fair if it is sold for $P(H) \times M$ dollars (Hájek, 2008: 173).

With this in mind, then “a Dutch Book is a set of bets bought or sold at such prices so as to guarantee a net loss” (Hájek, 2008: 174). So, if you regard a bet fair, and if entering this bet means that you will be guaranteed to a sure loss no matter how the world turns out to be, it can be said that you are Dutch Bookable. Moreover, in Dutch Book arguments, the bookie (a person who offers you the bets) and the agent who enters the bet are required to have the same knowledge that is relevant for the betting scenario. In other words, the bookie is required to not know anything that the person who enters the bet does not know. Under such conditions, if an agent has a set of credences and also a

Dutch book can be made against these credences, we can say that such openness to Dutch Bookability signifies incoherence in the person's (set of) belief(s). It should be noted that, the sole purpose of the bookie is to win money off an irrational agent who has a set of incoherent credences (Hájek, 2008: 174). Dutch Books, thus, can be argued to be powerful way of modeling and analyzing degrees of beliefs and probabilistic systems.

The Dutch Book arguments are of two types: synchronic Dutch Book arguments in which the bets are made at the same time and diachronic Dutch Book arguments in which the bets are made at different times (Hájek, 2008: 185). I will now discuss some examples of Dutch Book arguments that illustrate probabilism.

To recapitulate, probabilism stands for the Bayesian view that the laws of probability apply to any credences (Hájek, 2008: 175). The laws or probabilities are axiomatized by Kolmogorov (as cited in (Hájek, 2008: 175)). These axioms are; (I) non-negativity which states that a credence cannot be negative, and (II) normalization which states that if X is a tautology then the credence for X is 1, and (III) finite additivity which states that probability of X or Y is equal to the sum of probability of X and probability of Y if X is independent of Y (Hájek, 2008: 175). I will now discuss Dutch Book arguments that apply to such axioms. But before, the reader should note that Dutch Book arguments work in the form of proof by contradiction. So, Dutch Books that are made for probability axioms start with the assumption that the agents violate them. This means that the agent regards a bet fair which is consistent with the violation of a probability axiom. The Dutch Book shows when such violation happens the agent will be subject to a sure loss.

Non-negativity: Here I will mention a Dutch Book Argument made for an agent who violates the non-negativity constraint. Such an agent is the one, who for a

proposition X , assigns a credence to X which is negative. In order to illustrate this, let us assume his credence for X is -0.5 . Consider the following bet:

If X holds, the bet pays 10 dollars. Otherwise, the bet pays nothing. According to the agent the fair price for this bet must be 10 times -0.5 that is -5 dollars. Then the agent should accept to sell this bet for -5 dollars. This fits with the agent's credence, but results in a net loss since selling for -5 dollars means the agent is going to pay 5 dollars for selling the ticket (Hájek, 2008: 176)

Normalization: In this case of violation of normalization, the agent assigns a credence less than 1 for a tautology X . Let us assume his credence for X is 0.6 . Consider the following bet:

If X holds, let us assume that the bet pays 10 dollars. But since the agent's credence in X is 0.6 he is ready to sell this bet for 0.6 times 10 dollars which yields 6 dollars. Since X is a tautology, it is certain that this bet will hold. It can be seen that the agent who sold his ticket for 6 dollars has lost $10-6$ which is 4 dollars. The normalization axiom can be violated if an agent gives the credence to a proposition which is bigger than 1. In this case let us assume that the agent's credence for some proposition X is 1.5 . In this case, the agent is willing to pay 15 dollars for a bet that pays 10 dollars if X holds, and nothing otherwise. Since X is a tautology, it is certain that X will hold. Hence the agent will win 10 dollars from a bet that cost him 15 dollars to enter. In this scenario the agent has lost 5 dollars (Hájek, 2008: 175).

Finite additivity: In the case of violation of finite additivity, the agent is willing to pay different prices to a bet on A and a bet on B , than a bet on A and B , when A and B are independent events. To illustrate the violation, the agent can be assumed to be willing to pay a price that is equal to 11 dollars for a bet on A and B . At the same time, if he is willing to pay 5 dollars for a bet on A , and

5 dollars for a bet on B, then he is violating finite additivity. His readiness to pay 5 dollars on a bet on A and 5 dollars to a bet on B also means that, if he holds the bets the agent is willing to sell such bets for the same prices. And if the agent sells this bets then he will make 10 dollars of income. But as we have assumed before, the agent is ready to take the bet on A and B for 11 dollars, when offered to him. Let us say that after he sells his ticket, the bookie offers him the bet that he is willing to pay 11 dollars for. Since the agent considers this bet a fair price, he will buy the bet on A and B for 11 dollars. But since he sold the bet on A and the bet on B for 5 dollars each, the agent has lost 1 dollar. If the agent was willing to pay 9 dollars, similar to the previous Dutch Book argument, the reverse of this strategy (buying instead of selling) would also be Dutch Bookable (Hájek, 2008: 175)

In addition to that, a Dutch Book for conditional probability can be made. Conditional probability of A given B is denoted as $P(A | B)$. The conditional probability can be calculated as (Hájek, 2008: 178):

$$P(A | B) = P(A \text{ and } B) / P(B) \text{ where } P(B) \text{ is greater than zero.}$$

Dutch Book argument for proving conditional probability involves conditional bets, proposed by de Finetti (cited in Hajek, 2009: 178). Conditional bet for A and B are;

- if A and B, pay 10 dollars.
- if not A and B, pay nothing.
- if not B, the bet is cancelled.

If the agent violates conditional probability rule, Hájek (2008: 175) argues that the agent would lose the bets. However I will not introduce the relevant bets

Hájek (2008) proposes which establishes the relevant Dutch Book arguments for conditionalization. I will not do so because such bets are not relevant for van Fraassen's (1984) diachronic Dutch Book argument which this thesis focuses on.

1.3 Diachronic Dutch Book Argument

Conditionalization rule applies to the case when the agent has some credence for X , and received new evidence E for X , then the agent updates his credence in accordance with E in the following way:

Let $P_{old}(X)$ stand for the initial credence for X . Then, $P_{new}(X) = P_{old}(X|E)$ where $P_{old}(E) > 0$ (Hájek, 2008: 175). Lewis (1999: 406) argues that it is rational only if an agent updates by conditionalization and should the agent use any other update rule than conditionalization, a diachronic Dutch Book argument can be made against her.

This accounts for the background about the Dutch Book arguments that I believe was necessary in order to provide the stage for van Fraassen's (1984) discussion of the Reflection Principle in the 'Belief and the Will'. Van Fraassen motivates the principle from a diachronic Dutch Strategy argument, for this reason I will now provide a discussion of such strategy.

1.4 Van Fraassen's Diachronic Dutch Strategy

Van Fraassen (1984: 240) argues that if a diachronic Dutch strategy can be made against an agent, then there are two possible reasons for this. The first reason is about the way the agent updates her beliefs. If the agent is updating her beliefs in a way that violates conditionalization rule then that makes her vulnerable to a Dutch strategy. The second reason concerns the initial beliefs,

not the way they are updated. If the initial beliefs are not coherent, then the agent can become vulnerable to a Dutch strategy. Van Fraassen (1984: 240) constructs a Dutch strategy regarding the second reason in order to analyze how the initial beliefs should be in line with each other.

Van Fraassen's Dutch strategy focuses on two propositions. Let us assume that, H is a hypothesis and E is about the agent's future attitude towards the hypothesis. P denotes the current and P_t denotes the future possibility (Van Fraassen, 1984: 240). The Dutch strategy has the following bets (Van Fraassen, 1984: 240) can be seen in Table 1.

Table 1: Van Fraassen's Diachronic Dutch Strategy

	Pay	Price	Condition
Bet 1	1	$P(\neg H \wedge E)$	If $\neg H \wedge E$
Bet 2	$P(\neg H E)$	$P(\neg H E) \times P(\neg E)$	If $\neg E$
Bet 3	$[P(\neg H E) - P_t(\neg H)]$	$[P(\neg H E) - P_t(\neg H)] \times P(E)$	If E

Van Fraassen (1984: 241) creates a scenario which involves an agent and a bookie who makes the Dutch Book strategy against the agent. Initially, the agent has degrees of belief for a hypothesis H . Then, the bookie asks the agent to consider a bet on H at odds two to one. The agent's degree of belief in E denotes the agent's confidence in the bet the bookie offers. The agent is also encouraged to guess his chance for assessing the hypothesis mistakenly. Then, in order to show numeric results, the agent's credences for such propositions are taken as $P(E) = 0.4$ and $P(\neg H \wedge E) = 0.3$. The bets are considered on the grounds that H will be $1/3$ which is denoted by E . By using these, following values can be computed:

- $P(\neg H|E) = P(\neg H \wedge E)/P(E) = 0.3/0.4 = 3/4$
- $P_t(\neg H) = 1 - P_t(H) = 1 - E = 2/3$

When these values are plugged in to the bets, the bets can be seen in Table 2:

Table 2: Computed Values for Bets

	Pay	Price	Condition
Bet 1	1	0.3	If $\neg H \wedge E$
Bet 2	0.75	$0.75 \times 0.6 = 0.45$	If $\neg E$
Bet 3	1/12	$1/12 \times 0.4 = 1/30$	If E

The total price of the bets is $0.3 + 0.45 + 1/30 = 0.75 + 1/30$. If $\neg E$, then the agent only wins the second bet and earns 0.75. Then his net loss would be $0.75 + 1/30 - 0.75 = 1/30$. If E , the third bet is won. Also, the bookie offers to buy a bet from the agent on the future possibility of H for $2/3$. Notice that this is a diachronic Dutch Book. Follow-up bets are essential in the diachronic Dutch Books. Now, the net gain is $2/3 + 1/12 = 0.75$. After subtracting the total price, again the agent has a net loss of $1/30$ (Van Fraassen, 1984: 241).

Comparing with the synchronic Dutch Book, no non-tautology has the probability equal to one and every proposition can be decided in a finite amount of time (Van Fraassen, 1984: 242). If $P_t(\neg H)$ were 0.75, the agent would not have any loss. If that was the case, then $P_t(\neg H)$ would be equal to $P(\neg H|E)$. According to Van Fraassen (1984: 242), this means that the agent's future probabilities for event X is not in line with the current probabilities for X which suggests the Reflection Principle as a rationality constraint for credences. The Reflection Principle is formulated by Van Fraassen (1984: 244) in the following way:

$$P_{t_0}(A|P_{t_1}(A) = r) = r$$

In this formulation, P_{t_0} denotes the credence function of an agent at some time t_0 , and P_{t_1} denotes the credence function of an agent at some later time t_1 . r is equal to the agent's degree of belief in proposition A at time t_1 . Thus, the principle requires that the agent's credence r in A in the future time t_1 , matches the agent's credence in A at the present time t_0 . In the rest of the thesis, I will

evaluate the plausibility of this principle.



CHAPTER 2

SLEEPING BEAUTY

In this chapter I will provide a discussion of the (rather popular) Sleeping Beauty Problem (Elga, 2000). Elga (2000) argues that the Sleeping Beauty Problem proposes a challenge to the Reflection Principle (van Fraassen, 1984). Briefly, in Elga's (2000) thought experiment, the agent is argued to be rationally required to update her beliefs in the face of no non self locating evidence. Elga proposes this case as a violation of the Reflection principle (van Fraassen, 1984) because the agent no longer endorses an agreement with her expected future credence in the same proposition now.

2.1 Sleeping Beauty Problem

Elga (2000: 143) formulates the Sleeping Beauty Problem (SBP) in the following way:

“Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you to back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree ought

you believe that the outcome of the coin toss is Heads?”

Elga (2000: 144), initially considers two different answers to the puzzle that he offers. The first answer he acknowledges is that Sleeping Beauty’s credences that the coin lands Heads should be $1/2$ (Elga 2000: 144). This answer seems to be intuitive, because our credence in a fair coin toss is $1/2$, which can be generalized into Beauty’s credences before the experiment. It also seems to be intuitive to argue that Beauty’s degree of belief in the coin toss should remain to be $1/2$ when she learns about the experiment she is about to participate, because nothing in the experiment as she learns, interferes with the fairness of the coin toss. So taking those into consideration, we may expect that when Beauty wakes up on Monday, she should continue to have the same degree of belief as she had before the experiment. The $1/2$ response is also supported because Beauty seems to have learned nothing new about the world while the experiment. For example, on Monday when Beauty wakes up she does not learn that all coins in the world now are like triangular prisms and thus three sided. If she learned such a fact then perhaps she would have a reason to update her prior belief while she is participating the experiment. A rational update of one’s belief would be allowed on the face of new evidence , and when there is no evidence it follows that there is no reason for updating the belief.

Notice here that there are no relevant worries about conditionalization based on evidence here for Beauty. Since the discussions about whether an agent should give the evidence credence 1 or something less and update by Jeffrey Conditionalization (as cited in Meachem 2015: 769)- which enables the agent to give credence to evidence based on a weighted average of mutually exclusive and weighted compatible propositions - (Meachem, 2015: 769), does not arise here for Beauty. This is because, this view just argues that Beauty receives no evidence in the first place. Therefore her credence in a fair coin toss landing

Heads should remain to be the same, just as before her participation in to the experiment (Elga, 2000: 144).

The second answer Elga (2000: 144) acknowledges is that her credence should be $1/3$. Elga's (2000) second answer seems to stem from a *frequentist* understanding of probabilities. The frequentist interpretation can be taken to be defining the probability of an event as being tantamount to the frequency limits which the outcome is displayed in long series of comparable events (Gillies, 2012: 2). In the Beauty's case, if the experiment would be repeated infinitely many times, and in each of those experiments Beauty gave the answer Heads, Beauty would be one out of three times correct (Elga, 2000: 144). For that reason Beauty's credence in the proposition that the coin lands Heads should be $1/3$ (Elga, 2000: 144).

2.2 Elga's Argument and It's Consequences for the Reflection Principle

Among the two answers Elga (2000: 144-145) considers, Elga (2000: 144-145) picks the one thirder position as the correct one. However in order to support his claim Elga (2000) formulates a different argument than the one presented above. The argument he gives elaborates on the temporal locations which are captured as the states in the experiment (Elga, 2000). In order to provide the setting for how Elga's (2000: 144-145) argument unfolds, let states be temporal locations that are captured in the experiments. Also let for each state X , X stands for the name of the state. Also let each state consist of a tuple of two parameters denoting the outcome of the coin toss, and the day of awakening. Then, these states are:

0. State X : $\langle \text{the outcome of coin toss, day of the awakening} \rangle$

1. State S1: < “heads”, day: Monday >
2. State S2: < “tails”, day: Monday >
3. State S3: < “tails”, day: Tuesday >

Taking this configurations and the corresponding states of the world in consideration, Elga (2000: 145) observes that S2 and S3 are different temporal locations of the same possible world. In other words, they happen to be true in a possible world where the result of the coin toss is tails (Elga, 2000: 145). To recapitulate, in the Sleeping Beauty problem, Beauty is asked to state her credence for the coin toss landing Heads when it is the first time of her awakening. This situation corresponds to S1, in the sense that Beauty is in S1. Thus, calculating $P(S1)$ where P is the function that maps the state into corresponding credence is essential for Beauty. For this Elga (2000: 145) claims that $P(S2)$ is equal to $P(S3)$ as they are of the same possible world, and $P(S1)$ is equal to $P(S2)$. He supports his claims that :

1. $P(S2)$ is equal to $P(S3)$, and
2. $P(S1)$ is equal to $P(S2)$ in the following way:

Elga (2000: 145) first asks us to assume that, after the awakening, Beauty learns that the outcome is Tails, then Beauty knows that she is either in S2 or S3 (since S1 and S3 is already defined to include the Tails outcome). Elga (2000: 145) argues that, in both temporal locations, there is not any comparable difference for her. The difference does not arise for Beauty because according to the experimental design Beauty can not tell what day it is. Therefore Elga (2000: 145) argues that Beauty should assign equivalent credences to $P(S2)$ and $P(S3)$. This deals with Elga’s (2000: 145) claim (1).

In addition to that, to show (2) (that $P(S1) = P(S2)$), Elga (2000: 145) adopts a different strategy. Elga (2000: 145) argues that during the experiment, the researchers might conduct the experiment in two following ways:

1. First toss the coin then wake Beauty up either once or twice based on the outcome (twice if Tails, once if Heads)
2. First wake her up once (on Monday), then toss the coin to determine if they will wake Beauty up again. If the coin lands Tails, they will wake her up again

Then Elga argues that should Beauty want to calculate her credence for Heads, Beauty can either regard the first option or the second. Notice here that Beauty is searching for states where the outcome of the coin toss is Heads. In the first case the coin has landed Heads and Beauty is woken up only once. On the other hand, in the second option since Beauty is woken up only once, it follows that the coin has landed Heads. In both cases, Beauty should come up with the same result. Since both options are indifferent, Elga (2000: 145) uses the second option. Furthermore, assuming that Beauty is awake, and Beauty learns that it is Monday, it is warranted for Beauty to think that she is in either S1 or S2 (Elga, 2000: 145). With that knowledge in mind, Beauty can conclude that what makes S1 and S2 different is the outcome of the coin toss, since in both states the day of awakening is Monday (Elga, 2000: 145). Also, again since Beauty knows that it is Monday, $P(S1)$ is equivalent to $P(\text{Heads})$. Similarly $P(S2)$ is equivalent to $P(\text{Tails})$. In this case (on Monday), Beauty's sample space contains coin landing Heads and coin landing Tails. Thus Beauty can conclude that $P(\text{Heads}) = 1/2$. As $P(\text{Heads})$ is equal to $P(S1)$, and we already assume that it is Monday, it can be said that $P(\text{Heads}) = P(S1 \mid \text{it is Monday})$ which is equal to $P(S1 \mid S1 \text{ or } S2) = 1/2$. Thus, $P(S1) = P(S2)$

(Elga, 2000: 146).

So, Elga (2000: 146) concludes that $P(S1) = P(S2) = P(S3)$. Since $P(S1) + P(S2) + P(S3) = 1$ (according to finite additivity), $P(S1)$ is equal to $1/3$.

According to him, this shows a change in Beauty's credence from $1/2$ to $1/3$ (Elga, 2000: 146). Before the experiment, Beauty has the credence of $1/2$ about the proposition of a fair coin toss landing Heads, and later her credence in the same proposition, Elga (2000: 146) argues ought to be $1/3$. Also, before the experiment Beauty was aware of all the rules (Elga, 2000: 146). However, Elga (2000: 146) argues that, there is a change in terms of the relevancy of one's temporal location to the truth of the proposition. Before the experiment, one's knowledge of temporal location does not matter (Elga, 2000: 146). On the other hand, Beauty's temporal location matters for Beauty during the experiment (Elga, 2000: 146).

In the light of this discussion, Elga (2000: 147) claims that Sleeping Beauty Problem produces a counterexample to the 'Reflection Principle' (as cited in Elga, 2000: 147). The violation happens because without this experiment, Beauty would assign $1/2$ to the proposition that the result of a fair coin is Heads. However throughout this experiment Beauty should assign $1/3$ to the same proposition. However the Reflection Principle demands one to assign the same credence to a proposition now that she thinks she will assign in the future, given that she received no new information (van Fraassen, 1984). Elga (2000) does not further explain how exactly the Reflection Principle (van Fraassen, 1984) is violated himself. So I will try to come up with a view that seems to be compatible with Elga's (2000) position.

So, if Beauty took this Reflection Principle (van Fraassen, 1984) seriously, she may be required to do two things. Let us assume that :

1. -B denotes Beauty just before she participates in the experiment
2. B denotes Beauty while the experiment
3. B+ denotes Beauty after the experiment

Hence, in one scenario, -B assigns $1/2$ to the proposition of a fair coin toss landing Heads, and learning or contemplating about this possible experiment, she realizes that there is a future Beauty, B who assigns $1/3$ to the very same proposition. In this case, -B updates her credence to $1/3$. However she has no reason to do that, thus -B should violate the Reflection Principle and have non matching credences (van Fraassen, 1984).

In the second case, if Beauty took the principle (van Fraassen, 1984) seriously, during the experiment she (B) may end up thinking that after the experiment she (B+) will assign $1/2$ to the proposition that a fair coin landing Heads. So this means that, according to the principle (van Fraassen, 1984), B assigns the credence $1/2$ to the proposition, during the experiment. However as Elga (2000) argues, Beauty should assign $1/3$ credence to the proposition that coin lands Heads during the experiment. Hence in both of the cases, Beauty is violating the Reflection Principle (van Fraassen, 1984). Elga (2000) argues that violation happens because B does not learn anything new that -B does not know. This same relation seems to also hold for B and B+.

So, it seems to be possible that any explanation that can show us that B learns something new that -B does not know seems to resolve the tension. If B learns something new, then B can rationally believe in the proposition of a fair coin toss landing Heads with the credence of $1/3$, while -B assigning $1/2$ to the very same proposition, without any Beauty violating the principle of Reflection. This conclusion is motivated because the Reflection Principle only demands agreements with the future self under the condition of the current self and

future self shares the same evidence base. The explanation that seems to be at stake here has been provided by Hitchcock (2004), which I will discuss in the next chapter as alternative explanations to the Sleeping Beauty Problem.



CHAPTER 3

ALTERNATIVE EXPLANATIONS

In this chapter, I will evaluate two alternative explanations of the Sleeping Beauty Problem; one of which is offered by Hitchcock (2004), and the other which is offered by Bradley and Leitgeb (2006).

Among various alternatives to Elga's (2000) answer in the literature, I have chosen to discuss Hitchcock's (2004) for two reasons. The first reason that I will use to justify the selection of Hitchcock's answer is related to him being a proponent of one thirder view, similar to Elga (2000). However, despite that, Hitchcock (2004) argues for a one thirder answer which is compatible with the Reflection Principle (van Fraassen, 1984) (Hitchcock, 2004). This breaks the link between being a one thirder and being against the Reflection Principle. Such result means that one thirder view may be compatible with the Reflection Principle.

The second reason of Hitchcock's selection is related to Hitchcock's usage of the diachronic Dutch Book arguments in order to support his conclusion, similar to van Fraassen (1984). I believe that Hitchcock's (2004) usage of the diachronic Dutch Book arguments as a support for his conclusion provides a natural way

to evaluate the success of the Dutch Book arguments in general. This seems to be important because the Reflection Principle (van Fraassen, 1984) that is central to this thesis is also motivated by a diachronic Dutch Book Argument. So any attack on the diachronic Dutch Book arguments may have a direct impact on the Reflection Principle itself. After all, we may be willing to reject a principle which we think arises from a motivation that is not a reasonable motivation in the first place.

Now, I will try to justify my selection of Bradley and Leitgeb's (2006) analysis among many others. The main reason for their selection is related to their conclusion that betting odds and credences may come apart. In order to show how this may be the case, interestingly, they use a Dutch Book argument as well. However, I have initially argued that my selection of Hitchcock (2004) analysis was dependent on my interest in evaluating the convincingness of the Dutch Book arguments in the first place. At this point, one may legitimately react against my justification here about my choice of including Bradley and Leitgeb (2006) view in the analysis! Afterall, Bradley and Leitgeb (2006) paper uses a Dutch Book argument to show that credences and fair betting odds come apart sounds self defeating in the first place. However, I would like to note here that their usage of the Dutch Book arguments are designed to show what appears to be a Dutch Book argument from Hitchcock's perspective may not be in fact a legitimate Dutch Book. This conclusion is supported because they argue that Beauty is Dutch Booked because of her unfortunate peculiar epistemic state given the experiment. However, if a reader still has doubts about my last argument, I believe that I can try to convince them with another one. If the reader still believes that a Dutch Book argument can not be used in order to reject Dutch Books, then I invite them to consider an analogous case where one event which we initially think to be a cause of something yields two incompatible results. In such cases, I think a reasonable thing to do is to look

for another cause to explain our results. Similarly if we think that Dutch Book arguments causes two incompatible results one in favor of the endorsement of the Reflection principle and the other against it, then I guess Bradley and Leitgeb paper will gives us a means to legitimately consider the Reflection Principle itself (without its connection to a Dutch Book argument) as a stand alone rationality constraint. This is what I will do in the last chapter of my thesis. However, before that I think a reasonable task would be to explain the alternative explanations of Hitchcock (2004) and Bradley and Leitgeb in the first place.

3.1 Self Locating Beliefs Count as Evidence

Hitchcock (2004: 405) argues that Sleeping Beauty Problem is different than usual problems of uncertainty. Not only that the outcome of the event, but also the position in the temporal locations within possible worlds are uncertain (Hitchcock, 2004: 405). Similar to Elga (as cited in Hitchcock, 2000: 408), Hitchcock (2004) analyzes two initial answers: $1/2$ and $1/3$ and he provides his own argument in support of the $1/3$ answer. He supports his claims with diachronic Dutch Book arguments for both of the cases.

For the one-third case, Hitchcock (2004: 406) first adopts a frequentist approach. Should the experiment be repeated many times, half of the coin tosses will result in Heads, and the other half will result in Tails (Hitchcock, 2004: 406). In case of Heads, Beauty will be woken up once and in the case of Tails, the Beauty will be woken up twice (Hitchcock, 2004: 406). Thus, the frequency of her awakening in the case of Heads will be $1/3$. Hitchcock (2004: 406) thinks that the frequentist argument has some flaws. He argues that, the frequentist approach is suitable for the relative frequencies when each trial is

independent (Hitchcock, 2004: 406). In the Sleeping Beauty Problem, the trials are Beauty's awakenings. However, when the coin lands Tails, twice awakening is warranted, and such trials are not independent (Hitchcock, 2004: 406).

Second awakening is dependent on the first awakening, when the result of the coin toss is Tails (Hitchcock, 2004: 406).

For the proponents of the $1/2$ answer, Beauty's prior probability is $1/2$ and she also knows all the rules of the experiment beforehand. During the experiment, she gains no new information, thus her credence should not change at all (Hitchcock, 2004: 406). Hitchcock (2004: 406) finds the $1/2$ case more reasonable on the grounds of probabilistic reasoning. Beauty did not receive any new information, so she should not change her credences and she did not change her credence. This conforms to the rule of updating beliefs by conditionalization which states that if the prior is $P(H)$, and E is learnt, the credence should be updated to $P(H|E)$. Similarly, the second case is in accordance with the Reflection Principle which states that if Beauty thinks that her credence for $P(H)$ is equal to X at time t_2 , she should also have the same credence for $P(H)$ at some earlier time t_1 , if she did not learn anything new (as cited in Hitchcock, 2004: 407). Furthermore, supported by probabilistic reasoning, Hitchcock (2004: 406) thinks that any option that counters the $1/2$ case violates the rule of updating credences as well as the Reflection Principle. According to his view (2004: 406), any violation of conditionalization will make a person prone to a diachronic Dutch Book written against her. Furthermore, he (2004: 407) claims that the Reflection Principle is justified by a diachronic Dutch Book.

Hitchcock (2004: 407) constructs a diachronic Dutch Book argument to show that if Beauty would update her belief from $1/2$ to anything else, she would be prone to a sure loss. Considering the one-third answer as a case for such a violation Hitchcock (2004: 407) offers two bets :

1. The first bet on Heads which costs 1.5 dollars to enter; if $P(\text{Heads})$ holds, pays 3 dollar. Else, pays nothing. This bet is offered before Beauty is going to sleep.
2. The second bet on Tails which costs 2 dollars to enter; if $P(\text{Tails})$ holds, pays 3 dollars. Else it pays nothing. This bet is offered after Beauty's awakening.

It can be seen that both bets will be bought by Beauty, since she thinks that her initial credence is $1/2$ she buys the first bet. Moreover, since her credence while the experiment becomes $1/3$, she will be judging the second bet as fair and correspondingly willing to enter the second bet (Hitchcock, 2004: 407).

Now, the outcome of the coin toss is either Heads or Tails, so Beauty earns only one bet. This corresponds to a gain of 3 dollars for her (Hitchcock, 2004: 407). However, she paid a total of 3.5 dollars for the tickets. No matter the outcome, her loss of .5 dollars is guaranteed. For any view that suggest Beauty to have a credence anything but $1/2$, a similar Dutch Book argument can be created. However, Hitchcock (2004: 407) claims that such Dutch Book argument is problematic. Before supporting this claim, he first offers his own argumentation for the $1/3$ answer (Hitchcock, 2004: 407).

Hitchcock (2004: 408) offers an altered version of Elga's (2000: 144-145) argument for the one-third view. Similar to Elga (2000: 144-145), Hitchcock's symmetry argument cares about the temporal location of Beauty. Let states be temporal locations that are captured in the experiments, each state consists of a name and a tuple of three parameters denoting the outcome of coin toss, the day of the week, and Beauty's present status. Then, such states will be of the following form :

State X: $\langle$ the outcome of coin toss, day of the week, status of Beauty $\rangle$

Then, Hitchcock talks about the credences over such states, and makes three relevant observations about them:

1. His first observation is that the passage of time does not affect the outcome of the coin's landing (Hitchcock, 2004: 408). Thus, for any day of the week, let δ denote the day, $P(<H, \delta, \text{"Awaken"}>)$ is equivalent to $P(<T, \delta, \text{"Awaken"}>)$. This means that at any day, $P(H) = P(T) = 1/2$.
2. His second observation concerns prior probabilities for the days of the week (Hitchcock 2004: 409) . What credence should Beauty have for $P(\text{Monday})$? Similar to previous calculation, $1/7$ (seven comes from the number of the days of a week) would be a fair prior (Hitchcock 2004: 409).
3. His third observation is the need to represent Beauty's conditional degrees of belief in compliance with her knowledge about the experimental setup (Hitchcock, 2004: 409). We know that, if it is Monday and the coin lands Tails, Beauty is awake. Let $P(A)$ to signify probability of being awake. Thus $P(A|\text{Monday} \wedge T)$ is equal to 1. Similarly $P(A|\text{Monday} \wedge H) = P(A|\text{Tuesday} \wedge T) = P(A|\text{Monday} \wedge T) = 1$.

Then, Hitchcock (2004: 409) evaluates the credences over such states. When Beauty is awake, she only knows that she is awake. Thus her credences become, for example, $P(\text{Monday} \wedge T|A)$ meaning that given she knows she is awake, she tries to figure out what is her credence for the day being Monday and the outcome is Tails (Hitchcock, 2004: 409). Similarly, with the information provided above,

$P(\text{Monday} \wedge T|A) = P(\text{Monday} \wedge H|A) = P(\text{Tuesday} \wedge T|A) = 1/3$. Thus, $P(H|A) = 1/3$ which means that Beauty's credence when she wakes up is $1/3$ (Hitchcock, 2004: 410).

Hitchcock (2004: 411) thinks that his symmetry argument for $1/3$ may help to explain the argument for $1/2$. Although argument for $1/2$ claims that nothing new is learnt, according to Hitchcock, Beauty learns that she is awake (Hitchcock, 2004: 411). What she learns is a centered proposition and differs from a regular proposition (Hitchcock, 2004: 411). Probabilistic reasoning accounts suggest that upon learning new information one should conditionalize the prior credences. If conditionalization upon learning new information has not occurred, then a diachronic Dutch Book can be written to show that it is the losing hand. In the same sense, Hitchcock(2004: 411) argues that upon learning new centered proposition, one should again update the prior credences and Dutch Book arguments can be written to the ones who violate this condition .

So considering his various arguments so far, Hitchcock (2004) argues that both $1/2$ and $1/3$ cases can be supported with diachronic Dutch Book arguments. This is paradoxical if it is not erroneous. One caveat is that, without offering a proof, he claimed that $1/2$ case is problematic. Hitchcock (2004) reveals why he thinks the initial case is problematic. The reason for this problem is that Bookie knows more than he should (Hitchcock, 2004: 412). Dutch Book arguments can not be established when the bookie and the agent do not share the same knowledge. However, Hitchcock (2004: 412) argues that the bookie in the first diachronic Dutch Book argument create bets with the information that is not present to Beauty. The discrepancy in the shared information between the bookie and Beauty is not about the knowledge of the outcome of the toss since like Beauty, bookie does not have this information (Hitchcock, 2004: 412). On the other hand it seems that bookie knows what day it is when Beauty wakes up, while Beauty lacks this information (Hitchcock, 2004: 412). Moreover, since Bookie does not experience the change in temporal location as Beauty does, his bet creating behaviors are different than Beauty's bet fairness evaluations (Hitchcock, 2004: 412). According to Hitchcock (2004: 412), one

way to ensure that there are no such discrepancies between the bookie and Beauty is to require that bookie and Beauty fall asleep and wake up at the same time. What follows from that is that, now the bookie who is not aware of which day of the week he is in, becomes unable to know if he placed the next bet or not. Thus, the bookie must offer the same bet when he wakes up. So, in this scenario, if the outcome of the coin toss is Tails, Beauty and the bookie will wake up two times. In this case the bookie will offer two bets (since he wakes up twice). On the other hand, if the outcome is Heads, they will wake up one time and the bookie will offer one bet.

Now, to recapitulate the claim we want to check in the first place was whether Beauty should remain to assign the same credence to the proposition that the fair coin toss landing Heads. So should she assign $1/2$ as she used to, while the experiment (Hitchcock, 2004: 413). In order to check Beauty's credence before the experiment the bookie can offer a bet on Tails to Beauty which costs 1.5 dollars to enter and pays off 3 dollars if Tails and nothing otherwise. Since prior to experiment Beauty's credences in Heads and correspondingly Tails is $1/2$, we expect her to judge this bet as fair and enter. Then, during the experiment when the bookie wakes up with Beauty, he can offer her a bet that costs 1 dollar to enter and pays off 2 dollars if Heads and nothing if not. As shown above, if the coin lands Tails, Beauty and the bookie will be woken up again and since bookie can not tell which day it is, he will offer the same bet which costs 1 dollar to enter and pays off 2 dollars if Heads and nothing if otherwise. If Beauty's credence is still $1/2$ then she will enter all of the bets above. So if the coin lands Heads, Beauty loses the before the experiment bet and wins the bet on Monday. In this condition Beauty pays both bets 2.5 dollars and wins only 2 dollars, she will be on .5 dollars loss. On the other hand, if the coin lands Tails, Beauty will win the prior to experiment bet, and lose two of the while the experiment bets. In this condition Beauty pays 3.5 dollars and wins 3,

so she loses .5 dollars. So no matter how the world turns out to be (independent of the result of the coin toss), Beauty is prone to a diachronic Dutch Book. So if Beauty does not bet as if her credence in Heads is $1/3$ she will be prone to an ensured loss.

3.2 Lessons from Hitchcock and Elga

Before proceeding, I believe that it is necessary to compare the morals of the stories about Sleeping Beauty as it is offered by Elga (2000) and Hitchcock (2004), in order to evaluate the Reflection Principle (van Fraassen, 1984) as a rationality constraint.

So to wrap up, I will argue that from the Sleeping Beauty problem, Elga (2000) concludes that Beauty learns nothing new about the world while she is in the experiment, that she has not learned before. Despite that, Elga (2000) believes that Beauty should change her credence in the proposition that the fair coin toss landing Heads to $1/3$ while the experiment. Elga (2000) believes that such change happens because in the experiment Beauty's temporal location becomes relevant in determining something about the world. Elga (2000) also admits that this still yields a violation to the Reflection Principle (van Fraassen, 1984) because the change of relevancy of an already known information does not count as having new evidence. Hence, Elga (2000) believes that rationality may require someone to be in disagreement with an imagined future epistemic self now.

However, I will argue that Hitchcock (2004) instead argues that the change in centered proposition counts as learning new information. Beauty learns about her being awake, when she wakes up. So at this point it is not clear why Hitchcock (2004) counts this as a case of acquiring new information, because

prior to the experiment Beauty already knows that she will be woken up. But, how is being woken up different from knowing that one will be woken up in the future is unclear and very puzzling to me. Hitchcock (2004) also believes that Beauty should assign the credence of $1/3$ for the proposition of the fair coin toss landing Heads, since if she does not she will be Dutch Booked. However, this view is compatible with the Reflection Principle (van Fraassen, 1984) because according to Hitchcock (2004) Beauty has learned something new, so she is allowed to change her belief. So, all things considered the violation of the Reflection Principle (van Fraassen, 1984) may still depend on whether Beauty really has learned something new or not. Moreover, Hitchcock (2004) may not have provided a strong reason for why being awakened at some t_2 has more information than knowing that one will be woken up at t_2 when it is t_1 (assuming that t_1 is some time before t_2). So the Reflection Principle (van Fraassen, 1984) better not depend on this. However, despite that Hitchcock also shows Beauty would be Dutch Booked if she would not be a one thirder. Hence, if it is still possible that Beauty may not have received new information, then her Dutch Bookability as a halver may provide grounds for rejecting the Reflection Principle (van Fraassen, 1984). The next move in this argument may be to check if the link between betting odds and credences is as strong as Hitchcock (2004) may wish them to be, because if it is not then perhaps Beauty does not need to update her credence and thus violate the Principle. The argument of the sort is proposed by Bradley and Leitgeb (2006).

3.3 Credences and Fair Betting Odds Coming Apart

Bradley and Leitgeb argue that it is wrong to require the rational agent to match her credences with their fair betting odds, in all of the cases (2006: 119). They argue that it is possible for rational agents to have credences that do not

track their fair betting odds (Bradley & Leitgeb, 2006: 119).

In the case of Sleeping Beauty, Bradley and Leitgeb (2006: 122) agree with Hitchcock (as cited in Bradley and Leitgeb, 2006: 122) that Sleeping Beauty is prone to being a Dutch Book written about her, if she does not enter into bets as if her credences were one third for the proposition of a fair coin toss landing Heads. However they also argue that, Beauty's unique epistemic case requires that her fair betting odds do not follow her credences (2006: 120). In order to support their conclusion Bradley and Leitgeb give examples of two cases; which they call (a) the forgery case and, (b) the hallucination case where the agent's fair betting odds are shown to come apart from their rational credences (2006: 120).

In the (a) forgery case, Bradley and Leitgeb (2006: 123) offer two betting scenarios. In the first betting scenario an agent is offered one, or no bets based on the result of a fair coin toss. For the ease of explanation of the betting scenario here, I shall name the person who is to be offered the bet Barbara. After the toss of the fair coin, if the outcome is Heads the bookie offers Barbara no bets (Bradley & Leitgeb, 2006). However, if the coin lands Tails the bookie offers a bet on Heads to Barbara, which is about the coin that is already tossed (Bradley & Leitgeb, 2006: 123). A brief summary of the betting scenario is provided in Table 3.

Table 3: Forgery Scenario 1

Outcome of the Coin Toss	Bet	Outcome
HEADS	NO BET <NO PAY OFF>	No Change
	BET (ON HEADS)	
TAILS	<Costs \$ 0.5> <Pays Off \$ 1>	– \$ 1

Bradley and Leitgeb (2006: 123) argue that Barbara should not enter such a

bet, since in this scenario there is no way that Barbara can win any money. Instead, if she enters the bet anyway she surely will lose her money (Bradley & Leitgeb, 2006: 123). This is because the bookie only offers the bet on Heads to her when it is in fact the case that the coin landed Tails (Bradley & Leitgeb, 2006: 123). However, given that, Bradley and Leitgeb argue that it is not yet possible to conclude that Barbara's credences and betting behaviour come apart in this scenario. After all the fact that Barbara is offered a bet can inform Barbara about the forgery. Being alarmed by the trickster bookie, on the basis of being offered a bet in the first place, Barbara can lower her credence in the proposition of a fair coin landing Heads (Bradley & Leitgeb, 2006: 123). Correspondingly, on the basis of being offered the bet Barbara can raise her credence in the proposition that the coin has landed Tails, almost up to a certainty level (Bradley & Leitgeb, 2006: 123). That kind of a result does not seem to be counter intuitive, since as rational agents we may have reasons to raise our credence in the negation of a proposition when the proposition is declared by those who we know to be liars. Similarly, when Barbara updates her credences and she will no longer be willing to take the bet on Heads, since she will not recognise the offered bet as fair (Bradley & Leitgeb, 2006: 123). Thus, Barbara's fair betting odds still track her credences (Bradley & Leitgeb, 2006: 123).

However, Bradley and Leitgeb offer another scenario where being offered a bet does not alarm Barbara that coin did not land Heads (Bradley & Leitgeb, 2006: 123). In this betting scenario, if result of the toss of a coin is Heads, the bookie offers a bet on Heads to Barbara. But unknown to both of them, their notes are changed with fake money, so when they bet on Heads they bet with fake money. In the case where the coin lands Tails, the bookie offers (Bradley & Leitgeb, 2006: 123) a bet on Heads to Barbara. For the ease of illustration I will assume that in both outcomes the cost to enter the bet for Barbara is (fake or real) 0.5

and the pay off is (fake or real) 1, which seems to be compatible with Bradley and Leitgeb's offer (2006). A brief summary of the betting scenario is provided in Table 4. Bradley and Leitgeb (2006: 123) then asks, whether Barbara should enter this bet and they give a negative answer. They argue that it would be irrational for Barbara to enter such bets because Barbara either does not win any real money or she loses some real money (Bradley & Leitgeb, 2006: 123). However they argue that Barbara's credence in the coin landing Heads should remain $1/2$ (Bradley & Leitgeb, 2006: 123). In this scenario, they argue to have made it the case for betting behaviours and credences coming apart (Bradley & Leitgeb, 2006: 123). They argue that the existence of a fake bet (in which the agent is unable to realize the fakeness of the bet) should not be a reason for guiding her credences (Bradley & Leitgeb, 2006: 123).

Table 4: Forgery Scenario 4

Outcome of the Coin Toss	Bet	Outcome
HEADS	BET (ON HEADS)	
	<Costs fake 0.5c>	fake \$ 1
	<Pays Off Fake 1 \$>	so no gain
TAILS	BET (ON HEADS)	
	<Costs real \$ 0.5>	-\$0.5
	<Pays Off real \$ 1>	

However one may still believe that such a conclusion is unwarranted. This conclusion is motivated by Barbara's being unaware of the unfairness of the actual situation. In the second betting scenario of the forgery case, given the conditions it seems to be compelling to think that Barbara may enter this bet despite a sure loss. What seems to be determining her behaviour in betting scenarios is that her credences match the fair betting prices. In this case it is just unfortunate that Barbara loses money, but the loss is due to conditions that are not known to her. It seems to be that, only an outsider with the knowledge of the real condition is in a position to tell that Barbara should not

enter this bets. However given her knowledge Barbara can enter this bet. So it is possible to argue that her credences should guide her betting behaviour.

On the other hand, I believe that it may be possible to dismiss such worries. I think that what Bradley and Leitgeb (2006), wants to argue here is not about what determines one's behavior. I think what they are disputing is not about whether the agents should act in accordance with their credences or not. But what seems to be at stake here is that betting behaviours should be guided by actual winning situations and not credences. Although as not idealized agents this option is not available to us. What is needed for Bradley and Leitgeb (2006), to make their point is the existence of such cases where the credences may be conceived as coming apart from betting behaviours.

Bradley and Leitgeb (2006: 124) offer another betting scenario which they call the (b) hallucination case. They argue that this is another scenario where betting behaviours and credences diverge (Bradley & Leitgeb, 2006: 124). The hallucination case seems to be interesting because they use this betting scenario as a comparative case to the Sleeping Beauty Problem (Bradley & Leitgeb, 2006: 124). In this betting scenario, for the ease of discussion lets call our agent Barbara again. This time the betting scenario has two stages analogous to Beauty's awakening days. In this scenario, if the coin lands Tails, in both stage 1 and 2 Barbara will be offered real bets on Heads (Bradley & Leitgeb, 2006: 124). If on the other hand, the coin lands Heads, Barbara will be offered one real, one hallucinatory bet on Heads (Bradley & Leitgeb, 2006: 124). Barbara can be offered a hallucinatory or real bet on stage 1 or stage 2 (Bradley & Leitgeb, 2006: 124). If she has the hallucinatory bet on stage 1, she will have the real bet in stage 2, whereas if she has the real bet on stage 1, she will have the hallucinatory bet on stage 2 (Bradley & Leitgeb, 2006: 124) The bets are all on Heads. This yields four bets (Bradley & Leitgeb, 2006: 124). A brief

summary of the betting scenario and the results is provided in Table 5. It can be seen that a combined outcome of the four bets yields a negative expected utility (Bradley & Leitgeb, 2006: 124). However the probability of a coin toss landing Heads remains $1/2$ since among all bets $2/4$ cases correspond to the outcome of Heads landing (Bradley & Leitgeb, 2006: 124). But the outcome of bets yield a negative expected utility (Bradley & Leitgeb, 2006: 124). So, if Barbara would enter such bet she is prone to a sure loss (Bradley & Leitgeb, 2006: 124). Also, she is unable to realise the unfairness of this bet, and related Dutch Bookability of herself because of her unique epistemic state (Bradley & Leitgeb, 2006: 124). Thus Bradley and Leitgeb (2006: 124) argue that this is an example where again betting odds and degrees of beliefs are no longer in line with each other.

Table 5: Hallucination Scenario

Outcome of the Coin Toss	Bets (on Heads)		Outcome		Expected utility of 4 bets
HEADS	BET A: Stage 1 $\vee$ Stage 2 Hallucinatory	BET B: Stage 1 $\vee$ Stage 2 Real	BET A: no gain or loss	Bet B: gain of (assume) one utility unit	$0+1-1-1$ $= -1$
TAILS	BET C: Stage 1: Real	BET D: Stage 2: Real	BET C: loss of (assume) one utility unit	BET C: loss of (assume) one utility unit	

As indicated before, Bradley and Leitgeb (2006: 125) argue that hallucination scenario is an appropriate model for analyzing Sleeping Beauty Problem. They argue that the applicability of the hallucination model to Sleeping Beauty Problem becomes evident when instead of hallucinatory bets we consider the agent as sleeping (Bradley & Leitgeb, 2006: 125). When the agent is sleeping she is offered no bets (Bradley & Leitgeb, 2006: 124). In addition to that, we

are also invited to consider that the agent in this condition is drugged so as to block her from knowing that it is the first or the second bet that she is being offered (Bradley & Leitgeb, 2006: 125). The appropriateness of the model unfolds in the way that; in the hallucination case, if Barbara is offered a real bet (on Heads) in stage 1, and if the outcome of the coin toss is Heads then she will be offered no bet in stage 2 (Bradley & Leitgeb, 2006: 125). However given that she is offered a real bet on Heads in stage 1, if the outcome of the toss is Tails, she will be offered a bet on Heads which is a losing bet (Bradley & Leitgeb, 2006: 125). Hence, Barbara should not enter the stage 2 bet. Similarly, Beauty is only offered a bet on Tuesday (on Heads) if the coin landed Tails (on Monday). Hence, Beauty should not enter the Tuesday bet on Heads, since it is the losing bet (Bradley & Leitgeb, 2006: 125). But due to the experimental conditions Barbara/Beauty does not know if she is offered the fair bet (the real bet on stage 1 / Monday bet), or the unfair bet (the bet in stage 2 when the coin landed Tails/ Tuesday bet) (Bradley & Leitgeb, 2006: 125). Barbara/Beauty could take the strategy of calculating the total expected utility of the whole bets only to conclude that the expected utility is negative (Bradley & Leitgeb, 2006: 125). A brief summary of the betting scenario and total expected utility is provided in Table 6.

Table 6: Sleeping Beauty Bets

Outcome of the Coin Toss	Bets (Days)	Expected Utility +1 signifies wins the bet -1 signifies loses the bet Bets are on the same stakes
HEADS	Monday	Bet on Heads if coin lands Heads = +1 if coin lands Tails = -1 Total expected utility for Monday bets = 0
TAILS	Monday	
HEADS	Tuesday NO BET	Bet on Heads if coin lands Heads = -1 Total expected utility = -1
TAILS	Tuesday	
		Total Utility = -1

Bradley and Leitgeb (2006: 125) argue that like Barbara, Beauty should not enter any bets. This result is motivated because, Beauty does not know what day it is which means that she does not know if she is being offered the fair bet on Monday or the unfair one on Tuesday (Bradley & Leitgeb, 2006: 125). Yet she knows that the expected utility of the bet is negative, and if she enters sets of bets with a negative utility someone can throw a Dutch Book at her (Bradley & Leitgeb, 2006: 125).

However, Bradley and Leitgeb (2006: 125) argue that to make it the case for Hitchcock (as cited in Bradley and Leitgeb, 2006: 125) type of Dutch Books, the bets should be considered as stand alone bets instead of as a whole package. As Bradley and Leitgeb (2006: 125) argue, Hitchcock (as cited in Bradley and Leitgeb, 2006: 125) argues that Beauty should enter each bet. Also, Hitchcock (as cited in Bradley and Leitgeb, 2006: 125) argues that Beauty's credence in the proposition that the coin landed Heads should be $1/3$ if she wants to avoid being Dutch Booked when she enters the each bet. Bradley and Leitgeb (2006: 125) argues that in the hallucination case, similarly Barbara can avoid being Dutch Booked if she enters the bets as if her credence in the proposition that the fair coin toss landed Heads equals to $1/3$. However they argue that this does not give a reason for Barbara to "believe that the probability of Heads is $1/3$ " (Bradley & Leitgeb, 2006: 125). Bradley and Leitgeb (2006: 125) argue that, since Barbara's case seems to be shown to be parallel to that of the Beauty, like Barbara, the fact that Beauty avoids being thrown a Dutch Book at her only if she entered each bet as if her credence in the proposition is $1/3$, does not give a reason for Beauty to "believe that the probability of Heads is $1/3$ ". They argue that Hitchcock (as cited in Bradley and Leitgeb, 2006: 125) can not escape this conclusion too. This is because of two reasons. They argue that Hitchcock (as cited in Bradley and Leitgeb, 2006: 125) does not give any reasons as to why; (1) the Beauty might have an information that Barbara

lacks, (2) Barbara's credence should be $1/3$ (Bradley and Leitgeb, 2006: 125).

Bradley and Leitgeb (2006: 126) argue that credences should not be guided by betting behaviors under cases where the bookie only offers a bet if it is the losing bet, and the agent who is being offered the bet by the bookie is unable to tell the unfairness of the bet due to her unique epistemic state. After all being offered a losing bet, which you are not able to track the unfairness of should not be a reason for your credence in the proposition of a fair coin toss landing Heads (Bradley and Leitgeb, 2006: 126).



CHAPTER 4

REJECTING THE REFLECTION PRINCIPLE

In this chapter I will evaluate the Reflection Principle by itself as a rationality constraint. By the end of this chapter, I will come to the conclusion that the Reflection Principle is not a convincing rationality constraint because it demands that epistemic agents give up their agency.

But before that, I want to evaluate the process of modeling rational agents since the very same process seems to provide grounds for the principles of rationality or constraints thereof, among which we may include van Fraassen's (1984) proposal. First of all, it seems to be reasonable to argue that when we are modeling the rational agent, we are relying on our intuitive conception of rationality. For example, we think that conditionalization by Kolmogorov's ratio analysis is a rational way for updating beliefs. But, it seems to be also the case that, in order to make this apply to actual agents, the epistemic agents are required to have neat epistemic states which corresponds to for example, the Kolmogorov ratio formula. However, at the same time, we also observe that the actual agents are messier than that. Thus, in order to account for that, we make various simplifying assumptions about beliefs (since they are at the focus for our modeling purposes), such assumptions, for example, include the

requirement that people do not give zero to the condition when they update their beliefs by conditionalization. We enforce such restrictions on beliefs because otherwise our ratio formula gives us no defined probability due to the problem of division by zero (Hájek, 2003) .

It looks like a similar process is taking place in van Fraassen's (1984) account as well. Van Fraassen's (1984) intuitive conception of rationality seems to require him to make a certain set of assumptions about the belief states of actual agents. At this point it seems to be relevant to first interpret what may be included in van Fraassen's (1984) intuitive conception of rationality. It could be argued that his conception of rationality includes two dimensions; (1) the epistemic integrity component, and (2) the commitment component.

About the (1) epistemic integrity component, it seems to be possible to argue that van Fraassen's (1984) rationality understanding is ensuring that; (1.1) the agent is not changing her beliefs on an arbitrary basis, such as based on no new evidence, and (1.2) that the rational agent is defined over the coherence of her belief now and her belief about a future belief. To elaborate on (1.2), it could be said that van Fraassen's (1984) rational epistemic agent can be defined to be an epistemic agent solely on the basis of her agreement between her subjective probability for P now, and her subjective probability P at some later time on the basis of no new evidence. The disagreement in that sense cannot account for the integrity of the epistemic agent, since beliefs are conceptualized as subjective probabilities.

In addition, it could be argued that van Fraassen's (1984) intuitive conception of rational beliefs encompasses a (2) commitment component, which seems to be the idea that beliefs are like promises. They are commitments to which, if they are endorsed, one should not direct a question like do you really believe that? Both of the speculations about what van Fraassen's (1984) intuitive

conception of rationality seems to be compatible with the Reflection Principle (after all he endorses voluntarism about beliefs).

Having investigated what van Fraassen's (1984) intuitive conception of rationality captures, it seems to be possible to investigate what kind of assumptions such a conception of rationality enforces on the epistemic states of (ideal) agents. First of all, it can be said that to satisfy such demands it seems reasonable to argue that van Fraassen requires that epistemic agents have belief functions that are complete. In other words, the epistemic agent is required to assign a corresponding degree of belief to any proposition. However, it seems to be also the case that such idealization does not correspond to actual agents, since actual agents do not assign a degree of belief to any proposition. Instead the actual agents seem to have probability gaps.

At this point it is important to clarify what probability gaps are. Probability gaps happen when a proposition "is assigned no probability value whatsoever" (Hájek, 2003: 278). On de Finetti (as cited in Hájek, 2003: 278) type interpretation of probabilities, probability gaps happen when one just has no credences assigned to propositions about a bet scenario (Hájek, 2003: 278). This can happen when an agent is not ready to enter a bet at the time of the offering (Hájek, 2003: 278).

Hájek argues that van Fraassen (as cited in Hájek, 2003: 278) type understanding of subjective probability, requires that, if one lacks the corresponding behavioral dispositions, then any credence attribution is allowed for an agent as long as it falls between the 0-1 interval, including the border values of 0 and 1 (2003: 278). But Hájek (2003: 278) also argues that such forced credence attribution does not reflect such agents' epistemic state. This is because we can imagine a person B, and a proposition R, where B may perceive R to depend on a honing of a skill, which requires time (Hájek, 2003: 278).

Following from that it seems to be possible to argue that for most actual agents probability gaps happen about propositions that require them to conditionalize on their future beliefs. And previously, I have argued that, van Fraassen (1984) instead requires that epistemic agents have complete credence functions. This idea was motivated because van Fraassen (1984) requires the agents to assign a subjective probability to a proposition now, considering an expected subjective probability assignment in the future, which seems to be not satisfied for most of the beliefs of actual agents. At this point one could object and argue that the Reflection Principle does not enforce a complete credence function on individuals, since it does not actually require that one conditionalize on each and every belief. Instead it seems to be possible to argue that the Reflection Principle asks agents who aspire to be rational, to only match their future expected credences with their credences now, if they have considered their future epistemic self and corresponding future subjective probabilities. Hence, it may be argued that the Reflection Principle's assumptions with regard to actual agents are harmless.

But I will argue that this very assumption that van Fraassen needs for theoretical purposes may not be as harmless a requirement that is expected to be fulfilled by the rationality aspiring actual agents as it may appear in the first glance.

First of all, the reasons for how what one expects her future self to be like, ought to have an effect on what she believes today seems to be unjustifiable. When one shapes her current credence on the basis of her expectations about her future self, one seems to be undercutting her agency. At this point, one could consider that a similar relation holds between free will and determinism. Furthermore, one could maintain that, since the tension between free will and determinism can be resolved by some version of a compatibility argument,

similarly the tension between the Reflection Principle and epistemic agency may be resolved in principle by a similar line of argument. However I think that such resolution may be possible for free will and determinism without it being applicable to the tension between epistemic agency and the Reflection Principle. For example, in the free will and determinism debate there seems to be a dichotomy between the individual's intentions and actions. A compatibility argument can be given since in a deterministic world the congruence of intentions and actions seems to be sufficient. However this option does not seem to be available for the tension between the Reflection Principle and epistemic agency since, a similar dichotomy seems to be lacking for the agent. In epistemic agency the congruence links are to be made between two beliefs, which both have the commitment component as well. So the compatibilist response available for the free will and determinism debate seems to not be relevant for epistemic agency and the Reflection Principle.

Ruling the similarity relation out, I will show how and why the Reflection Principle violates epistemic agency. It seems to be possible to argue that the Reflection Principle violates epistemic agency because it requires that the epistemic agent attribute the same degree of belief to a proposition now which she thinks her future self will assign. This means that the agent has to regard the same evidences or set of beliefs as relevant in every case. This enforces that what the agent believes to be relevant features now will be carried over to her future attributions, or what she believes may be relevant evidence in the future is carried over for her current assessment. However belief formation may be context dependent. Agents who are to form a belief seems to be tracking such contextual factors, and doing exactly that seems to be a practice of epistemic agency. In the remaining part of the discussion I will try to illustrate how relevance links may be relevant for belief formation in the light of the discussions about the Sleeping Beauty Puzzle.

So, in the light of the Sleeping Beauty Problem, it has been shown in the discussion above that an acceptance of the Principle of Reflection requires two options:

1. To adopt the Bradley and Leitgeb (2006) view and argue that fair betting odds and credences may come apart. Since this view argues that Beauty has no reason to update her Belief the conclusion is that Principle is not violated by Beauty. This option does not require the revision that I will propose. However I think that this view is still compatible with the revision that I will propose.
2. To adopt the update view similar to Elga and Hitchcock
 - (a) The Hitchcock (2004) view argues that updating Beauty learns new information and hence does not violate the Reflection Principle.
 - (b) The Elga (2000) view argues that updating Beauty does not learn new information and thus violates the Principle.

However I have also argued that Hitchcock does not provide any substantial argument in favour of why learning self location counts as new information. One way to try to rescue the Reflection Principle might be to weaken its conditions on learning new information. So the revised version of the Principle may argue that it might be the case that what is learned is not new information, but a relevance link among Beauty's evidences and priors. Thus, the Reflection Principle's condition of "no new information is learnt" can be weakened to the following reformulation: "no new relevance link is made". In more detail, Let R be the set of relevance links based on an agent's evidences and credence for a certain proposition. Then for a proposition A , the agent will have credence function P and her credence is $P(A)$. Also, she has a set of information denoted by $E = \{e_1, e_2, \dots, e_n\}$. This information can be regarded as evidence. Then,

R_A denotes the set of relevance links between proposition A and some evidences that is member of E . For example, R_A might be equivalent to $\{e_1\}$ which means that this agent considers e_1 as the only evidence she has for the proposition A .

In the Sleeping Beauty Problem, A is the outcome of coin toss to be Heads. Before the experiment starts and Beauty is not given the rules of the experiment, Beauty's evidence set might be $E = \{\text{"the coin is fair"}, \text{"for a fair coin toss, both heads and tails outcomes are equally likely"}, \text{"For every action, there is an equal and opposite re-action"}, \dots\}$. Then, her relevance links might be $R_A = \{\text{"the coin is fair"}, \text{"for a fair coin toss, both heads and tails outcomes are equally likely"}\}$.

After she learns about the experiment setup, she updates her evidence set to be $E = \{\text{"the coin is fair"}, \text{"for a fair coin toss, both heads and tails outcomes are equally likely"}, \text{"For every action, there is an equal and opposite re-action"}, \text{"Sleeping Beauty Experiment Rule \#1"}, \text{"Sleeping Beauty Experiment Rule \#2"}, \text{etc.}\}$. Simply, she added the experiment rules to her evidence/ knowledge base. However, having such new information added to her E , does not mean that she updated her relevance set R_A at the exact time she received the evidences. It might be the case that she will never update the relevance set R_A during the experiment, thus the problem is then that she learns some new information, but she does not relate what is learnt with the proposition A . So the updating Beauty does not violate the revised version of the Reflection Principle because she has not updated her belief without new evidence, but her update can be accounted for by the new relevance link that she has made during the experiment, which is broken after the experiment. And doing exactly that seems to be a practice of epistemic agency.

REFERENCES

- Bradley, D., & Leitgeb, H. (2006). When betting odds and credences come apart: More worries for dutch book arguments. *Analysis*, 66(290), 119–127.
- Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis*, 60(266), 143–147.
- Gillies, D. (2012). *Philosophical theories of probability*. London: Routledge.
- Hájek, A. (2003). What conditional probability could not be. *Synthese*, 137(3), 273–323.
- Hájek, A. (2008). Dutch Book Arguments. In P. Anand, P. Pattanaik, & C. Puppe (Eds.), *The Handbook of Rational and Social Choice*, (pp. 173–196). New York: Oxford University Press.
- Hitchcock, C. (2004). Beauty and the bets. *Synthese*, 139(3), 405–420.
- Lewis, D. (1999). *Papers in metaphysics and epistemology* (Vol. 2). New York: Cambridge University Press.
- Meacham, C. J. (2015). Understanding conditionalization. *Canadian Journal of Philosophy*, 45(5–6), 767–797.
- Ramsey, F. P. (1931). Truth and Probability. In R.B. Braithwaite (Eds.), *The Foundations of Mathematics and Other Logical Essays*, (pp. 156–198). London: Kegan, Paul, Trench, Trubner & Co.
- Van Fraassen, C. (1984). Belief and the will. *The Journal of Philosophy*, 81(5), 235–256.