universität wien

Masterarbeit

Titel der Masterarbeit

Bayesian Epistemology of Disagreement

Verfasser

Patrick Lang, BA

Angestrebter akademischer Grad

Master of Arts (MA)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A066 941
Studienrichtung lt. Studienblatt:	Masterstudium Philosophie
Betreuer:	Mag. Dr. habil. Sven Bernecker, Ph.D.

universität wien

Master Thesis

Title

Bayesian Epistemology of Disagreement

Author

Patrick Lang, BA

Submitted in partial fulfillment of the requirements for the degree of

Master of Arts (MA)

Wien, 2014

Program ID number:	A066 941
Program:	Master of Arts in Philosophy
Advisor:	Mag. Dr. habil. Sven Bernecker, Ph.D.

For my mother

Contents

Acknowledgments

Several people supported me while I worked on my master thesis and I could not have completed it without them. On the one hand, I had support from my fiancée at home. Thanks for that, Maria. On the other hand, I received great support from my colleagues at the University of Vienna. In particular, I have to thank the Dean of the Faculty of Philosophy, ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth, who was always understanding and cooperative regarding questions of time management. At the end of each meeting, she also took the time to chat with me about my progress, which motivated me further. She deserves nothing less than my deepest gratitude. Last, but definitely not least, I must mention my supervisor, Mag. Dr. habil. Sven Bernecker. He not only consented to continue supervising me during my master thesis after moving back to the University of California, Irvine, but he did so in the best possible way. Whenever we had a Skype meeting, he spoke with me in detail about my progress and was always open to new ideas. I cannot imagine how I would have managed without him. Mr. Bernecker, thank you for all your support, the helpful inputs, and the hours of insightful discussions.

1

Introduction

Over the last decade, discussions around the topic of disagreement have become increasingly frequent. The field has created its own cases of interest—namely, cases of peer disagreement— and developed different ways to handle those cases. That all this proved to be more than just a flash in the pan is evident in that it even came to be named—the Epistemology of Disagreement (EoD). Some recent publications focus explicitly on this topic. The special 2009 issue of *Episteme* 6 (3) carried the topic, *The Epistemology of Disagreement*. 2010 witnessed the publication of *Disagreement*, a book edited by Richard Feldman and Ted A. Warfield. The latest publication—*The Epistemology of Disagreement: New Essays* by David Christensen and Jennifer Lackey—was published in 2013, and shows that the discussion is still ongoing. All this interest should not be surprising since it is—besides other kind of disagreements—the possibility of rationally believing something in the field philosophy itself that is being questioned.[1] This means that EoD raises the question of whether we as philosophers can at all have rational beliefs since we disagree about nearly everything. This reveals an important aspect of the discussion because, despite the actuality of the field of EoD, at its heart, it implies questions that have been known since ancient skepticism.[2]

In this work, I will attempt to combine this new field in epistemology with another, that of Bayesian Epistemology. The reason for this is that in EoD, nearly everyone speaks in terms of degrees of belief.[3] Since Bayesianism is a way to handle exactly those

[1] This for example, is the topic of Van Inwagen (2010).
[2] See Kelly (2005: 169) where he refers to Sextus Empiricus (2000) for some of the ideas that are discussed in the field of EoD.
[3] This happens for a reason, as I will show below.

degrees of belief and their dynamics, it seems more than appropriate to at least examine the end to which such a combination could lead. Another reason is that most of the works dealing with EoD talk about disagreement as a kind of information and ask how this information should be handled. In trying to answer this question, however, they only use cases to motivate sometimes this view and sometimes that view, and I have to admit, the cases are always very convincing. But since those different positions now for a decade confront each other with different cases, those cases seem to be mere *intuition pump*s.[4] Of course, everybody has different kinds of intuitions about disagreement. On the one hand, we want to be steadfast if we are in a disagreement about something that we believe to be important. On the other hand, we have the intuition that someone must change his doxastic attitude in the light of disagreement, especially if this someone is not us but our opponent. I do not know if this is the reason why the cases work in both directions, but I do have those intuitional moments when thinking about disagreement. Therefore, I thought it would be a good approach to consider the idea that disagreement is a kind of information, to find out what information it is exactly, and to handle this information in a formal way. In the end, I have to deal with the results of a formal analysis, which is hard enough even without turning over rival intuitions in one's mind.

I start *chapter 2* with a short overview of the field of the EoD. This chapter will introduce the main terms and define those cases in which EoD is interested. I will also present examples of how one can deal with and classify these cases.

In *chapter 3*, I will do the same for the concept of degrees of belief and Bayesian epistemology. I will present the basic concept of Bayesian epistemology and show how the approaches presented in this chapter influence the definitions given in chapter 2.

Since the concept of a defeater is an important one in the EoD, in *chapter 4*, I will attempt a formal analysis of cases involving defeaters. This will lead to some problems because the concept is defined in terms of a three-fold scheme of belief and a translation is not easy to find. Therefore, I will directly look at concrete cases and show how a Bayesian account can handle them.

[4] I borrowed the term from Dennett (1995).

The analysis of chapter 4 will go hand in hand with the one in *chapter 5*, since I will do exactly the same thing for higher-order evidence (HOE). The idea is to demonstrate that HOE is not different in principle from one kind of defeater—namely, the undercutting defeater. I will illustrate this by analyzing cases of HOE. Such an analysis is important because, besides a formulation of disagreement in terms of the defeater, a formulation in terms of HOE is very common. Therefore, this analysis is a preparation for the analysis of cases of disagreement.

This analysis of cases of disagreement will be carried out in *chapter 6*. Here, I will use the tool introduced in the previous chapters to describe the information that disagreement is giving us, and then to show how we can handle this kind of information. At the end of this chapter, I will visualize the impact of this analysis not only in cases of disagreement between cognitive equals but also on the interpretation of disagreement itself.

Finally, *chapter 7* presents the summary. This chapter will do more than just sum up the details. Here, I will provide a perspective on topics that I could not discuss in this work but where further investigations are necessary.

2

Epistemology of Disagreement (EoD)

Disagreement is not something that philosophers came up with in order to have a new topic of discussion. Disagreement is a common phenomenon in our everyday experience. The examples are countless: opinions about bedtime between parents and children, expectations about the results of a sports game between fans of opposing teams, a cat-lover and a dog-lover arguing over which pet is the best, the district attorney and the lawyer of a defendant, and so on. Equally countless are the common reactions of people to these situations. But EoD is not interested in the descriptive question of how people respond to a disagreement. The concern of EoD is the normative question of whether or not we should revise our beliefs in the light of peer disagreement, and if the answer to this question is a positive one, how we should do this.[5]

In this chapter, I will provide a short introduction to the field of EoD. The aim is to present and classify the basic possibilities for replying to the normative questions of EoD that have been roughly presented above. In 2.1, I will begin by specifying what disagreement is and, especially, what peer disagreement is, since this is the kind of disagreement relevant to our debate. A better understanding of the main problem of the field will provide a better perspective on the questions that EoD tries to answer. Thereafter, in 2.2, I will go on to present possible responses to these questions. Here, I will illustrate two basic distinctions by creating a 2*2 matrix that can categorize those answers. I will present five views with a short motivation for each and show how those views are related to the introduced distinctions.

[5] This normative question leads to several other questions like, does it matter what topic the disagreement is about, does it matter who was on the right track in the beginning, and so on. For a short outline of the topic and the questions discussed by others who have published in the field of EoD, see Frances (2010b).

The purpose of this chapter is not to argue for some of those views. The idea is to use those views, on the one hand, to show what positions are present in the EoD and, on the other hand, to introduce and work with the two important distinctions of the field.

2.1 Peer Disagreement

Disagreement is a common phenomenon in our everyday experience. Parents disagree with their child about bedtime, students disagree with their teacher about grades, a doctor disagrees with her colleague about the right treatment, and there are countless other examples. All of these cases have in common that one party believes some proposition p, while the other believes the negation of this proposition $\neg p$ or something that logically includes $\neg p$. For example, take the case of the bedtime disagreement. The parents believe that p is correct, where p stands for the proposition that it is time for the child to go to bed. Since the child thinks that it is not time for bed, it believes $\neg p$. Hence, the two parties disagree about the proposition p. In the case of the treatment disagreement, one doctor believes p, that the treatment should be some medication A. Her colleague thinks that the right treatment is medication B, and, therefore, believes q. Since this case should be an example for disagreement, we can assume that medication B rules out medication A; therefore, we get the relation $q \rightarrow \neg p$. Let us further assume that the doctor who believes in medication B knows about this relation, and also concludes on the basis of the beliefs q and $q \rightarrow \neg p$ that $\neg p$ is correct. Hence, this is again a case of disagreement about the proposition p.[6]

The assumption of knowing of each other's differing beliefs is implied by talking about disagreement. When we say that the parents and the child have a disagreement, we are thinking of a situation where the child reveals that she is unhappy about the decision of the parents. It seems related to consider this assumption when talking about disagreement. As Lackey (2010b) puts it, "it is not even clear that such a case [where opponents are not aware that they believe in the negation of each other's belief] properly

[6] The grade disagreement is analogous to the treatment disagreement, since one grade rules out the other.

involves a disagreement, let alone one that should be at the centre of this discussion" (303). Therefore, when I refer hereafter to disagreement, I have the following definition in mind.

D.2.1 (Disagreement): Two agents A and B disagree about some proposition p if and only if,

(1) A (or B) believes that p is correct,

(2) B (or A) believes that q is correct and is aware that $q \rightarrow \neg p$ (this is trivial in the case where $q = \neg p$) or is agnostic concerning p, and

(3) A and B are aware that they hold differing beliefs concerning p.

All three examples above are cases of disagreement in this sense. Besides this similarity, however, one must also note the difference. If we look at the bedtime disagreement and the treatment disagreement, we can see that the relations between the disagreeing parties differ. Sure, both parties are disagreeing, but in one case we have parents disagreeing with their child on a topic in which children are not really approved experts. It seems obvious that one should not give much weight to the child's belief that it is not bedtime. The situation in the case of the treatment disagreement is nothing like this. Here, we have two doctors, two approved experts about medical treatments who are disagreeing about the method of treatment. Due to this difference, the way out of the disagreement in the first case is an easy one, at least from an epistemic point of view. The child should listen to his parents since they are the experts in deciding what time is the right time to go to bed. The more complicated situation and therefore the more interesting one is the treatment disagreement. What should one doctor do, if she becomes aware of some other doctor who disagrees with her regarding the treatment of a patient? To exclude shortcuts for the doctor, we have to say something more about the disagreeing opponent. Therefore, let us further assume that he is not just a doctor who took a brief look at the files of the patient to formulate his belief. Both doctors have the same information about the case: the illness in question, and every other important detail. The fact that both doctors individually are equally good, trained experts would be insufficient for excluding such

shortcuts as in the bedtime disagreement. With this in mind, we can introduce another important definition to declare the main problem of EoD. The relation between the doctors that we wanted to express is the relation of two so-called peers, where peerhood has the following definition.

D.2.2 (Peerhood)[7]: Two agents A and B are peers relative to the question whether p if and only if they satisfy the following two conditions:

"*Evidential equality:* A and B are evidential equals relative to the question whether p when A and B are equally familiar with the evidence and arguments that bear on the question whether p." (Lackey 2010b: 302)

"*Cognitive equality:* A and B are cognitive equals relative to the question whether p when A and B are equally competent, intelligent, and fair-minded in their assessment of the evidence and arguments that bear on the question whether p." (ibid.)[8]

Before defining peer disagreement, we need some further clarifications concerning the beliefs of the opponents of a disagreement. To see why, let us take another look at a treatment disagreement. Now, it is not the doctor's colleague who disagrees with her but the patient herself, a five-year-old girl named Cindy. No one would blame the doctor if she did not take Cindy's objections seriously. But Cindy is not an ordinary little girl: she is a genius. Everyone is unaware that Cindy has studied everything that she could find about medicine in the last few days and is now as competent as the doctor concerning the question of her treatment. Cindy is following D.2.2, the doctor's peer, in the disagreement about her treatment. One could be interested in this sort of case, but since

[7] This definition follows Lackey (2010a: 274) and Lackey (2010b: 302). But similar definitions can be found in nearly every paper on EoD—for example, in Kelly (2005), from where I borrowed the term *peerhood*. I decided to follow Lackey because she names the two conditions of disagreement, and since this distinction is important below, it will be easier to refer to them by their names.

[8] Lackey also includes a third condition (*full disclosure*) in which she follows Feldman (2006). This condition claims that A and B shared their relevant evidence and arguments. One could argue that *evidential equality* also entails this condition as Ridder (2013: 2) does, but I think that this would not work. The difference is that with *full disclosure*, A and B are also aware of being evidentially equal, without needing to be aware of this fact in order to satisfy the condition of *evidential equality* alone. But I think that full disclosure can be substituted by evidential equality with a further condition that I will introduce in D.2.3.

they are not in the focus of the debate in EoD, the following definition of peer disagreement excludes such cases.[9]

D.2.3 (Peer Disagreement): Two Agents A and B are in the situation of peer disagreement concerning whether p if and only if

(1) A and B *disagree* about p

(2) A and B are *peers* concerning whether p is true

(3) A and B are *aware that* they are peers concerning whether p[10] is true

Obviously, condition (1) assures us that A and B are in a disagreement situation. But because of how we defined disagreement in definition D.2.1, specifically by point (3) of D.2.1, condition (1) in D.2.3 also assures us that A and B are aware of their disagreement. Following definition D.2.2, condition (2) says that A and B satisfy *evidential equality* as well as *cognitive equality*. The last condition excludes cases like the one between the doctor and Cindy.

2.2 Possible Answers to Peer Disagreement

The definition D.2.3 of peer disagreement from the last section gives us the exact outlines of those cases in which EoD is interested. Thus, we can now introduce a case that satisfies D.2.3 and use it to provide a concrete example of the normative question that EoD tries to answer. The case for discussion is the following disagreement between Mr. Spock and Scotty.[11]

[9] The same point is made by Lackey (2010b: 303–4).

[10] As mentioned above, *evidential equality* alone cannot do the job for full disclosure. But I believe that *evidential equality* and this *aware-of* condition can, since A and B are now as aware that they share the same evidence and arguments as is the case after *full disclosure*.

[11] Here and in the following, I will use for all the cases, characters from the original *Star Trek* show. I will describe these characters roughly following the show, but to evaluate a case one has to consider nothing more than the characterisation that is explicitly mentioned in the description.

C.2.1 (Warp Drive): Following an attack, the warp drive of the USS Enterprise has stopped working. Since the rest of the engineer corps died during the attack, Mr. Spock is now in the engine room to help Montgomery "Scotty" Scott. Even if Scotty is the chief engineer on board the Enterprise, they are both aware that Mr. Spock knows about the warp drive as much as Scotty does. Having evaluated all the relevant facts about the damage together, they are now discussing the possibility of fixing the warp drive. But although they have equal knowledge about the damage, and one is as good as the other in evaluating the possibility to fix the warp drive, Mr. Spock thinks that it is possible to fix it, while Scotty thinks that it is not.

It is easy to see that this is a case of disagreement in terms of D.2.1. Let us say that p_{WD} stands for the proposition that the warp drive can be fixed. Thus, we can say that Mr. Spock believes p_{WD}, while Scotty believes $\neg p_{WD}$. Therefore, case C.2.1 satisfies the first two conditions of D.2.1. Since Mr. Spock and Scotty are actually talking about whether p_{WD} is correct, they are also aware of each other's opinion. This means that Mr. Spock and Scotty also satisfy the third condition of D.2.1 and, therefore, the first condition of D.2.3. Case C.2.1 also describes the relation between Mr. Spock and Scotty as one in which both are in an evidential and cognitive equality; therefore, they are peers in terms of D.2.2.[12] Since they are also aware that they are each other's peers, the case satisfies the last two conditions of D.2.3 and is a case of peer disagreement. The concrete question that interests EoD in this case is what Mr. Spock and Scotty should do, given that they are in a peer disagreement. Let me now present answers that one can find in the literature of EoD. I will also introduce two distinctions that classify those responses to the question. The beginning makes the distinction between *conformists* and *nonconformists*.[13] After introducing this distinction, I will first discuss the conformist view (CV) and illustrate it using a well-known example. Thereafter, I will discuss different approaches for a nonconformist view (NCV). The difference between those approaches will also lead to the introduction of the second important distinction by demonstrating views that

[12] I am aware that describing Mr. Spock and Scotty as peers is not something to which everyone would agree. But let me clarify that first, I described them as peers only with respect to the question of whether p_{WD} is correct, and second, I will only refer to the explicit characterisation in the case description, which is just a rough sketch of the characterisation of the show.

[13] This is a distinction used by Lackey (2010a and 2010b).

accept rational disagreement and those that do not. I will call the former group the *rational disagreement view* (RDV) and the latter the *non-rational disagreement view* (NRDV). These two distinctions lead us to the following 2*2 matrix in which the different approaches can be placed side by side to provide an overview of their relationship.

CV + RDV	CV + NRDV
NCV + RDV	NCV + NRDV

The name of the CV already informs us that such a view will argue that Mr. Spock and Scotty have to adjust their beliefs with respect to each other, while an NCV thinks that this is not necessary. Following Lackey (2010b), we get the definition of the CV and the NCV that formulates this idea exactly.

D.2.4 (CV and NCV): Let us say that two agents A and B are in the situation of peer disagreement about a proposition p.

The CV says that A and B will have to bring their doxastic attitude towards p in line; therefore, they cannot rationally continue the belief/disbelief in p and have to become agnostic concerning the question of whether p is correct.

The NCV is a view that is not a CV.[14]

2.2.1 The Equal Weight View

The EWV was introduced in the EoD with this name by Elga (2007).[15] But the idea has a long history and had already played a prominent role in ancient skepticism and in the

[14] See Lackey (2010b: 299–300), where she presents this definition. She does not explicitly introduce the idea that A and B have to get their doxastic attitudes in line, but this seems to be the basic idea behind a CV. We will also see this in the example of a CV, which is presented as an example by Lackey.

[15] Christensen (2007) also defends a similar position without explicitly labelling it as EWV. But he drew a connection in Christensen (2009a), where he refers to Elga's EWV.

work of Sextus Empiricus.[16] The idea of the EWV can be formulated by the following definition.

D.2.5 (EWV): Let us say that two agents A and B are in the situation of peer disagreement about a proposition p.

The EWV says that A has to give B's belief the same weight as her own and vice versa.[17]

Advocates of the EWV often present cases like C.2.1 to motivate the EWV.[18] They do so because such cases seem to possess the intuitional power to motivate the EWV. Every aspect in these cases is designed to establish symmetry; thus, this seems to be the only reasonable weighting for the disagreeing beliefs as well. Even opponents of the EWV acknowledge this kind of intuitive motivational force and admit that in cases like C.2.1 a response like the EWV seems to be "almost trivial or obviously true" (Kelly 2010: 113). Pressing the intuition pump can be a good starting point, but not the final argumentation for a detailed view.

A detailed argumentation can be found in Van Inwagen (2010). Van Inwagen is actually trying to show the problems of a view like the EWV. Before starting with this, however, he offers a strong argument for such a view in order to make it obvious that if this approach leads to problems, they cannot be easily ignored. [19] He too starts with a case of peer disagreement, but goes on to bring out the assumptions on which such a view is built. His case of disagreement is the one between him and David Lewis. Van Inwagen points out that he is disagreeing with David Lewis about topics like free will, determinism and unrealized possibilities. A short examination demonstrates that what

[16] This historical reference can be found in Kelly (2005: 169). He refers to Sextus Empiricus (2000) and a position similar to the EWV, but since he published two years before Elga (2007), without the name. Nevertheless, he already used the phrase "give equal weight" (Kelly 2005: 13, 15, 29).
[17] A similar explicit definition can be found in Bogardus (2012). But his definition is, like mine, just a representation of the implicit definitions that one can find in the papers of the advocates of the EWV, such as Elga and Christensen.
[18] See Elga's motivation for the EWV using a case of perceptual disagreement between two peers (Elga 2007: 486–8), or Christensen's (2007: 193) famous restaurant check case.
[19] He uses this argument already in Van Inwagen (1996). Here, I follow the argumentation in Van Inwagen (2010: 23–4) where he uses this argument again.

Van Inwagen tries to establish with this case is a situation of peer disagreement as defined in D.2.3; therefore, his case is similar to C.2.1.

Van Inwagen considers himself to be an incompatibilist: he believes that free will is incompatible with determinism. Let us call the proposition that incompatibilism is true as p_{incomp}. David Lewis, on the other hand, thinks of himself as a compatibilist, which includes that he believes $\neg p_{incomp}$. Since Van Inwagen and David Lewis are both aware of this, we have a disagreement in the sense of D.2.1. This disagreement satisfies the first condition of a peer disagreement. Van Inwagen also says that he and David Lewis are aware of all the philosophical considerations relevant to p_{incomp}, and that the reason for the disagreement is not because one of them has some cognitive deficiency relevant to the question of whether p_{incomp} is true. Therefore, we can count them as evidential and cognitive equals, which means that they are peers with respect to this topic; hence, they satisfy the second condition of D.2.3. Now, given that Van Inwagen is saying all this and he has discussed it with David Lewis, we can say that both are aware of their peerhood. Therefore, this is a situation of peer disagreement over the proposition p_{incomp}.

As mentioned above, this is only the beginning of Van Inwagen's argument. The next step is to show what kind of thesis stands behind the assumption that disagreement is at all a problem. Van Inwagen shows that if the conclusion should be, that in cases of peer disagreement at least one of the opponents has to be irrational, then such a conclusion has to be based on the following thesis.

D.2.6 (Uniqueness Thesis):[20] The uniqueness thesis (UT) says that a body of evidence only justifies one doxastic attitude concerning a proposition p.

The UT together with a case of peer disagreement leads to the result that at least one of the opponents has to be irrational, since the definition D.2.3 of peer disagreement includes the evidential equality condition. In other words, peer disagreement is a situation with only one body of evidence which, by the means of UT, leads to the conclusion that

[20] UT was introduced and endorsed under this label by Feldmann (2007). But given that Christensen, for example, defends a version of the EWV, it should be no surprise the he has sympathy for such a view and adopts UT with some negligible changes under the name "rational uniqueness" (Christensen 2007).

under evidential equality, only one doxastic attitude is rational for both parties. But since we are talking about a disagreement, we have two doxastic attitudes; therefore, at the most, only one person can be justified in her doxastic attitude. In the terms of the warp drive case, this means that it cannot be the case that Mr. Spock and Scotty are justified in their doxastic attitudes.

But the UT alone would not lead to the EWV because it does not say that Spock *and* Scotty are irrational in their doxastic attitudes. In a semi-formal way, we could say that

(2.1) $\left(UT \wedge EVD_{p_{WD}}(Spock, Scotty)\right) \rightarrow ('B_{Spock}(p_{WD}) = IR' \vee {}'B_{Scotty}(\neg p_{WD}) = IR')$,

where $EVD_p(A, B)$ stands for two agents A and B having a disagreement under the condition of evidential equality towards some proposition p, $B_A(p)$ represents that agent A believes p, and IR stands for "irrational". To get from here to the result that both should be agnostic towards p_{WD}, which simply means that both are irrational in holding on to p_{WD}, we have to take a closer look at what irrationality and, therefore, rationality means.

D.2.7 (Epistemic Rationality): A is epistemically rational in holding the belief in some proposition p if and only if A's body of evidence favors p over $\neg p$.[21]

This definition of epistemic rationality includes as a corollary that if an agent A is in the situation that A's body of evidence favors p as well as $\neg p$, A must become agnostic toward p to be epistemically rational. Although this definition alone will not directly get us from 2.1 to the EWV, it will expose further the implication of this view. If peer disagreement as in C.2.1 or between Van Inwagen and Lewis really leads to agnosticism, it means that such cases are situations where this is the only way for the agents to be rational. Then, peer disagreement has to be a situation where both doxastic attitudes are favored equally by the body of evidence. This can be argued using the symmetry of the

[21] See Sosa (2010: 278), where he implicitly introduces such a definition.

peerhood. Since one peer is as likely to be right as the other, the opinions are as likely to be true.[22] Thus, we can rewrite 2.1 and see that the EWV has to argue that

(2.2) $\left(UT \wedge PD_{p_{WD}}(Spock, Scotty)\right) \rightarrow ('B_{Spock}(p_{WD}) = IR'\ ALA\ 'B_{Scotty}(\neg p_{WD}) = IR')$.

Here, $PD_p(A, B)$ stands for the two agents A and B having a peer disagreement towards some proposition p, and ALA stands for '*as likely to be true as*'. But the corollary of definition D.2.7 informs us that in order to be rational, one has to be agnostic in situations where one option is as likely as the other; therefore, we come from 2.2 to

(2.3) $('B_{Spock}(p_{WD}) = IR'\ ALA\ 'B_{Scotty}(\neg p_{WD}) = IR') \rightarrow ('B_{Spock}(p_{WD}) = IR' \wedge\ 'B_{Scotty}(\neg p_{WD}) = IR')$.

On summarizing, we get that after the disagreement, Scotty and Mr. Spock have a situation that favors neither p_{WD} nor $\neg p_{WD}$, ; therefore, both have to become agnostic regarding whether p_{WD} is true. The same holds for Van Inwagen and David Lewis disagreeing over whether p_{incomp} is true. So, given our definition D.2.3 of peer disagreement, together with the UT from D.2.6 and the definition of epistemic rationality from D.2.7, the EWV seems to be the right way to revise one's doxastic attitude in the face of peer disagreement. Objections to this answer come from advocates of the NCV, as we will see in the following section.

2.2.2 The Nonconformist View

As we have now seen the requirements on which the EWV is built, we have also seen possible ways to argue against this view. The first requirement for the EWV was the UT; without it, the argument would not get started. But the UT alone does not bring us all the way to the EWV. Therefore, we needed the idea that leads to 2.2: that both parties have a body of evidence, which does not favor one side over the other. The literature offers two

[22] For example, see Elga, who says, "[e]ven if in fact you have done a much better job than your friend at evaluating the claims, simply comparing your verdicts to those of your friend gives you no evidence that this is so" (Elga 2007: 487).

views that argue against this later foundation of the EWV. These views are the *correct reasoning view* (CRV)[23] and the *anti-symmetry view* (ASV).

Both, the CRV and the ASV, argue that in the light of peer disagreement one is justified in giving his own view some extra weight; thus, the symbolization 2.2 has to fail. As argued above, what leads to 2.2 is the idea of symmetry in the situation of peer disagreements. The idea behind this extra weight is now that even in cases of peer disagreement, something places one in an epistemic position that is superior to that of the other: a *symmetry-breaker*.[24] Both views accept the UT and argue that at least one position is not justified in the light of evidential equality. But while the CRV argues that only one position is arrived at by correct reasoning, the ASV denies the evidential equality. Here, we get the following definition of the ASV.

D.2.8 (ASV): A nonconformist response to peer disagreement counts as an *anti-symmetry view* if and only if the response argues against the possibility of evidential equality.

Approaches to an ASV can be found in Van Inwagen (1996 and 2010), and Sosa (2010). As mentioned above, Van Inwagen's motivations for a NCV are his concerns about a CV. After offering an argument for a CV, he begins to show the consequences of such a view. He reasons that because disagreeing with one's peer is such a common situation, this view would include that one has to be agnostic about nearly everything. Van Inwagen's response to this consequence is that he is not just unwilling, but also unable do to so, and he believes that this holds for nearly everyone else (cf. 2010: 28).

A way out of this misery, he opines, could be an account of evidence as something that cannot be totally shared between two persons. Van Inwagen provides examples to support this account of evidence. One of these examples is the famous *chicken sexer* case. Since only female chickens are useful in the production of eggs, the male chickens get sorted out and killed. Chicken sexers are the people who do the selection. The interesting thing is that although they are fast and highly reliable in their selection process, they

[23] See Elga (2007: 485) and Lackey (2010b: 302).
[24] I follow here the definition of symmetry breakers provided by Lackey (2010b: 39) who credits Nathan Christiansen in a footnote for coming up with it.

cannot pinpoint the exact evidence for their choices. Such examples seem to show that evidence is not necessarily exportable, and is at least partly private (cf. Van Inwagen 2010: 25–6). Consequently, Van Inwagen can still think of David Lewis, his opponent in the case of peer disagreement that he is discussing, as his cognitive equal, but deny evidential equality. The absence of evidential equality indicates that the UT does not apply here and Van Inwagen does not have to become agnostic about the question of whether p_{incomp} or $\neg p_{incomp}$ is true.[25]

If NCV does not want to deny the evidential equality and also take the UT in account, then the consequence is that at least one of the parties has to be irrational. This method leads us to the CRV.

D.2.9 (CRV): A nonconformist response to peer disagreement counts as a *correct reasoning view* if and only if the response argues that one is justified in giving his own belief extra weight in the face of peer disagreement because the belief in question is the product of correct reasoning (cf. Lackey 2010b: 300).

A formulation of a CRV can be found in the following quote from Kelly (2005), where he says:

> The rationality of parties engaged in [peer disagreement] will typically depend on who has in fact correctly evaluated the available evidence and who has not. If you and I have access to the same body of evidence but draw different conclusions, which one of us is being more reasonable (if either) will typically depend on which of the different conclusions (if either) is in fact better supported by that body of evidence.
>
> (Kelly 2005:180)

Here, Kelly's argument for the CRV relies on the UT. Since both parties have the same body of evidence as established by the evidential equality, and since UT says that only one doxastic attitude can be justified with this body of evidence, one party has to be irrational. So far, the CRV is in line with the EWV, which means that advocates of the CRV would agree with 2.1. The difference is that Kelly argues that only one of the parties should revise her doxastic attitude, while the other should remain unchanged. To be the

[25] An analogous argumentation can be found in Sosa (2010: 290–1)

one who got it right is an argument for holding on to one's doxastic attitude, while the EWV argues that because we do not know who got it right, this is no reason to do so. Therefore, the CRV is denying 2.2. But since 2.3 is a consequence of 2.2, the CRV is denying that both parties are irrational and therefore have to reverse their beliefs.

Both of these views, the ASV and the CRV, are different from the EWV and both have in common that they are accepting the UT. This implies that they also refuse rational peer disagreement, what means that it would be possible that in case of peer disagreement both parties are rational. It is true that the ASC from Van Inwagen argues that the disagreement between him and David Lewis is a rational one in the sense that both parties are rational in holding on to their beliefs. But they are rational in doing so because their own body of evidence favors their beliefs. However, there is no rational disagreement in cases of 'real' peer disagreement, where we have only one body of evidence for both parties, since the UT applies in those cases. By accepting the UT, and thereby denying rational disagreement, we have the second distinction that built the above-presented 2*2 matrix. Therefore, we must now talk about positions that are attacking the EWV by accepting rational disagreement.[26] Two views that are presented in the literature are the *egocentric view* (EV) [27] and the *total evidence view* (TEV)[28].

The EV says that one's own beliefs carry some extra weight because they are one's own beliefs. Lackey (2010b: 299) presents the following quote from Wedgwood as an example for the EV.[29]

> Perhaps, quite generally, it is rational for one to place greater trust in one's own intuitions, simply because these intuitions are one's own, than in the intuitions of other people. In other words, perhaps it is rational for each of us to have an egocentric epistemic bias in favour of our own intuitions.
>
> (Wedgwood 2007: 261)

From such a formulation of the EV, we can extrapolate the following definition.

---

[26] See Kelly (2005: 179–80) as well as Kelly (2010: 116), where he is arguing for the difference between those views that allow rational disagreement and those that do not.
[27] See Elga (2007: 485) and Lackey (2010b: 302).
[28] To be precise, Kelly (2010), who presents a formulation of the TEV, argues that this position can be formulated with or without the acceptance of rational disagreement. But here I will present the TEV as a position that accepts rational disagreement, which is also the position that Kelly himself favors (Kelly 2010: 116–7).
[29] Elga also refers to Feldman (2007: 209–10) where a similar view is presented, but is not endorsed by Feldmann.

D.2.10 (EV): A nonconformist response to peer disagreement counts as an *egocentric view* if and only if the response argues that one is justified in giving his own belief extra weight in the face of peer disagreement because the belief in question is his own belief (cf. Lackey 2010b: 300).

In the terms of C.2.1, we can say that if the EV is true, Scotty should trust his own opinion more than Mr. Spock's and therefore stand by his belief. But since the reason for this is that the belief is his own, and this is also true for Mr. Spock, the latter too can hold on to his belief. Therefore, it follows that the EV has to deny the UT. The EV does not argue that the opponents of peer disagreement have different bodies of evidence. Still, both can be reliable in holding on to different doxastic attitudes.

Let me now move on to the TEV as discussed by Kelly. Kelly presents the CRV but does not endorse it. In fact, he argues just as advocates of the EWV do against the CRV. He says that even if one's response to the body of evidence is the correct one, the fact that one's peer responds as he does should have an impact on one's doxastic attitude (cf. Kelly 2010: 136). His motivation for this objection to the CRV is a case of reiterated peer disagreement like the following one.

C.2.2 (Reiterated Peer Disagreement):[30] As we all know and as all Vulcans know, Vulcans and even half-Vulcans as Mr. Spock are all equally great logicians. But even in the light of this greatness, the Vulcan society has a logical conjecture, called *the Conjecture* (C), that has no proof. During his studies on Vulcan, the home planet of the Vulcans, Mr. Spock discovered a proof for C, which he believes is a sound proof. Let us call the proposition that Mr. Spock's proof is a sound proof $p_{C-sound}$. But after his important discovery he shows the proof to a Vulcan who studies the proof with great care, but reaches the conclusion that this proof is unsound. Due to the disagreement with this Vulcan, Mr. Spock shows the proof to more Vulcans. But all the Vulcans to whom Mr. Spock showed his proof independently reached the result that this proof is not sound.

[30] I present this case analogously to the case by Kelly (2010: 137) but will transform it into the Star Trek universe.

What impact should this series of peer disagreements have on Mr. Spock's belief in $p_{C-sound}$?

Kelly says that it is obvious that Mr. Spock cannot rationally hold on to his doxastic attitude for $p_{C-sound}$, even if the proof is in fact a sound one. He argues that even if the judgments of the Vulcans are misleading evidence, misleading evidence is evidence nonetheless. One must also consider this as an argument that the peer disagreement between Mr. Spock and his first Vulcan peer should have some impact on Mr. Spock's belief. If he were to give the first disagreement no weight at all, his situation in the second disagreement would be just the same as after discovering the proof. This would continue into the third, the fourth, and subsequent disagreements. If all these disagreements were given zero weight, then this would hold for the situation of the reiterated peer disagreement as a whole, which seems wrong (cf. Kelly 2010: 137–8). Thus, Kelly forms the following conclusion.

> One should give some weight to one's peer's opinion, even when from the God's-eye point of view one has evaluated the evidence correctly and he has not. But why? Exactly because one does not occupy the God's-eye point of view with respect to the question of who has evaluated the evidence correctly and who has not.
>
> (Kelly 2010: 138)

The question now is where the difference between the TEV and the EWV lies. Kelly attempts to bring together the right ideas of the CRV and the EWV. He continues to hold–likethe EWV does–that the opinion of the peer has some weight. He talks of peer disagreement as *higher-order evidence* (HOE). HOE is provides evidence about my first-order evidence (FOE) and how one should respond to one's FOE. But the difference between the TEV and the EWV is that, as per Kelly, the HOE counts for everything in the EWV. He also includes the idea of the CRV that one's FOE also has some weight. Using this idea, we formulate a definition for the TEV as follows.

D.2.11 (TEV): A nonconformist response to peer disagreement counts as a *total evidence view* if and only if the response argues that the right response depends on one's original FOE as well as on one's HOE presented by the peer disagreement.

The TEV counts as a NCV because the result of one's FOE together with one's HOE need not make such a huge change in one's total evidence as the CV thinks. In some cases, one's HOE will be strong enough that, together with the FOE, it will compel one to become agnostic. In other situations, however, one's FOE will be strong enough, or one's HOE will be too weak; as a result, it will not be the case that one has to become agnostic. According to definition D.2.4, the NCV is defined as a view, which states that one does not have to become agnostic but not that one must not become agnostic. Therefore, the TEV counts as a NCV. The TEV also includes the possibility of rational disagreement, since it does not rule out that, in some situations, the total evidence of both parties supports two different doxastic attitudes.[31] Therefore, the TEV also counts as a RDV.

In the following section I will summarize all these views and place them in the matrix of possible answers.

2.3 Classification of the Possible Answers

We have now seen five different responses to the situation of peer disagreement. I will now briefly summarize these views and classify them in the light of the two distinctions we have seen. Those distinctions are the CV-NCV, a distinction about the outcome of answers, and the RDV-NRDV, a distinction about the possibility of rational disagreement.

The EWV says that because the two opponents of a peer disagreement have the same HOE—namely, the peer disagreement itself—they also have the same possibility to be right, as was demonstrated in 2.2. Therefore, they have to give each other's doxastic attitudes the same weight and reach a joint solution.[32]

The NSV accepts the UT and thinks that only one party can be right in cases of peer disagreement. But in order to explain situations of disagreement with a person one

[31] Kelly says that the TEV neither excludes nor includes rational disagreement (2010: 116).
[32] As we have seen so far, the only possibility for such a joint solution is agnosticism, since if one peer believes p while the other $\neg p$ and both have the same possibility to be true, the only way to stay rational is to become agnostic about p.

considers a conditional equal, this view uses an account of evidence as at least partly private. This view is an account of the NCV and the NRDV.

The CRV thinks that what really counts is one's FOE. If one has evaluated this evidence right, he can hold on to his doxastic attitude. This view does not think that both parties should bring their doxastic attitudes closer together; therefore, this view is no CV. But it assumes that only one party is right, thereby excluding rational disagreement and, in doing so, implies the NRDV.

The EV thinks that one's own beliefs have some extra weight because they are one's own. Therefore, one can hold on to his beliefs even in situations of peer disagreement because peer disagreement is never a situation of symmetry, as definition D.2.3 seems to suggest. So, one does not have to become agnostic about the proposition of disagreement and, consequently, this view is a NCV. But since both parties can give their own view extra weight, it is also a view that accepts rational disagreement.

The TEV is a combination of the ideas of the EWV and the CRV. It takes both, the FOE and the HOE, into account. The result is a view that need not think that both positions must get the same weight; it can also make room for the possibility of rational disagreement. Therefore, the TEV is an account of the NCV as well as of the RDV.

Hence, these five views fit into the matrix built up by the two distinctions of interest in the following way.

	RDV	NRDV
CV	———	EWV
NCV	EV / TEV	CRV / NSV

As we can see, there is no possible view that combines the RDV and the CDV. This is because the CV excludes the RDV. The RDV says that one body of evidence need not support only one doxastic attitude. It does, however, not say that one body of evidence must not do so. But the RDV implies the possibility that, in some cases, both peers of a disagreement can be rational in their different opinions. But the idea behind the CV is

that both peers should give each other's beliefs the same weight and, therefore, should reach a joint solution. This means that the CV excludes situations where the opponents retain different doxastic attitudes even after the peer disagreement. Hence, there is no possible view in this field of the matrix.

3

Bayesian Epistemology (BE)

In epistemology, we have two different approaches to beliefs. One considers belief as a three-fold scheme, which is the account that was used in the previous chapter without explicitly mentioning it. Here, one either believes some proposition p, its negation $\neg p$, or is agnostic concerning p. But we can also think about belief as level of confidence; then, we have gradual differences of confidence towards some proposition p. This means that one is more or less confident that some proposition p is true.

In this chapter, I will show why the latter approach is preferred in EoD. Then, I will present a formal definition of probabilities using the Kolmogorov axiomatization. Using the so-called *Dutch book argument*, I will explain why such a definition should be considered as the condition for a rational degree of belief. Besides this rationality condition, I will also present the *principal principle*, which creates a relation between objective probabilities and one's rational level of confidence, and the *Bayesian conditionalization*, which is an account to render the rationality of the dynamics of beliefs. Since such a dynamic of beliefs is also of interest for EoD, the Bayesian conditionalization will be important in the following chapters. At the end of this chapter, I will take another look at what was said in chapter 2, while considering the new approach involving doxastic attitudes. I will rewrite previously provided definitions wherever necessary, because the change to degrees of belief has an impact.

3.1 Why Bayesian Epistemology?

In the chapter above, I spoke of belief as a three-fold scheme. Such an account of belief can be defined as follows.

D.3.1 (Epistemology of Beliefs):[33] Let $DA_A(p)$ be the doxastic attitude of some subject A concerning a proposition p, and let PDA_{EB} be the set of all possible doxastic attitudes. In case of an account of *epistemology of belief*, the following holds:

$DA_A(p) \in PDA_{EB}$, and

$DA_{EoB} \coloneqq \{belief, disbelief, agnostic\}$.

But this is not the only way to think about one's beliefs. Another approach sees beliefs, not as a three-fold scheme, but as the level of confidence. Here the possible doxastic attitudes are not the set PDA_{EB} with $|PDA_{EB}| = 3$, but represent subjective probabilities. This means that one's level of confidence represents the likelihood that one ascribes to some proposition being true. In the next section, I will present a formal axiomatization of such probabilities, but here I will define such an approach of belief as one, where we have a set of possible attitudes with a cardinal number that is not 3 but $|\mathbb{R}|$. Using this idea, we get the following definition of such an account of degrees of belief.

D.3.2 (Epistemology of Degrees of Beliefs): Let $DA_A(p)$ be the doxastic attitude of some subject A concerning a proposition p, and let be PDA_{EDB} the set of all possible doxastic attitudes. In case of an account of *epistemology degrees of belief*, the following holds:

$DA_A(p) \in PDA_{EDB}$, and

$PDA_{EDB} \coloneqq [0,1]$.[34]

[33] I have borrowed the distinction between *epistemology of belief* and *epistemology of degrees of belief* from Foley (1992).

In case of peer disagreement, the latter account of beliefs seems to be the appropriate one. Let us see why by changing the warp drive case C.2.1 slightly.[35]

C.3.1 (Warp Drive II): Following an attack, the warp drive of the USS Enterprise has stopped working. Since the rest of the engineer corps died during the attack, Mr. Spock is now in the engine room to help Montgomery "Scotty" Scott. Although Scotty is the chief engineer on board the Enterprise, they are both aware that Mr. Spock knows about the warp drive as much as he does. Having evaluated all the relevant facts about the damage, they are discussing the possibility to fix the warp drive. But although they have equal knowledge about the damage, and one is as good as the other in evaluating the possibility to fix the warp drive, Mr. Spock thinks that it is possible, while Scotty is agnostic about it.

As above, p_{WD} stands for the proposition that the warp drive can be fixed. If we follow D.3.1, we can say that the doxastic attitude of Spock concerning p_{WD} is $DA_{Spcok}(p_{WD})$ and that of Scotty is $DA_{Scotty}(p_{WD})$. Thus, the situation described in C.3.1 can be easily formulated using the following two equations.

(3.1) $DA_{Spock}(p_{WD}) = belief$

(3.2) $DA_{Scotty}(p_{WD}) = agnostic$

Since 3.1 and 3.2 still satisfy the conditions of the definition D.2.1, we have a disagreement about the proposition p_{WD} between Mr. Spock and Scotty, which can now be formulated as

(3.3) $DA_{Spock}(p_{WD}) \neq DA_{Scotty}(p_{WD})$.

Furthermore, since nothing besides $DA_{Scotty}(p_{WD})$ has changed between C.2.1 and C.3.1, we still have the situation of peer disagreement. But the slight change shows that the restriction to three possible doxastic attitudes leads to the result that the solution of such a case has fewer possible answers than the similar situation of C.2.1.

[34] Here, I do not define what the level of confidence means, or, in other words, when we should have which level of confidence. This will be presented later.
[35] I borrowed this approach from Kelly (2010: 117–8) to argue for an account of degrees of belief.

If we agree that $disbelief$ is not a possible solution for both opponents, then only the set $\{belief, agnostic\}$ is left. Therefore, we only have two principle answers for the situation of peer disagreement. Either Mr. Spock and Scotty give each other no weight at all and hold on to their doxastic attitudes or the opinion of just one of them has no weight at all and the solution is the doxastic attitude of the other. Views that would create a settlement between these two extremes are no longer possible, because there is no doxastic attitude between $belief$ and $agnostic$.

As an example, take a look at the EWV. In definition D.2.5, we saw that following the EWV means that Mr. Spock and Scotty should give each other's opinion the same weight as their own. This definition excludes both possible answers of the case C.3.1. If we take another look at the warp drive case of C.2.1, we can formulate the situation there similarly to the formulation of C.3.1 and get

(3.4) $DA_{Spock}(p_{WD}) = belief$

(3.5) $DA_{Scotty}(p_{WD}) = disbelief$

A solution that denies rational disagreement but does not only take one peer's opinion into account—such as the EWV—would settle somewhere between the doxastic attitudes of both peers. In a situation as described by 3.4 and 3.5, we would therefore get the following way out.

(3.6) $DA'_{Spock}(p_{WD}) = DA'_{Scotty}(p_{WD}) = agnostic.$

Here, $DA'_{Spock}(p_{WD})$ and $DA'_{Scotty}(p_{WD})$ stand for the doxastic attitude after the solution of the disagreement. Such a solution is not possible in a situation like C.3.1 where the doxastic attitudes are as in 3.1 and 3.2 because there is no doxastic attitude between $belief$ and $agnostic$.

Besides impacting the EWV, this also makes it impossible to get fine-grained versions of all the views above. For example, consider the EV. This is a position where one thinks that I can give my own beliefs some extra weight, because they are my own. But extra weight does not mean that I have to take my position into account. Such a view could be formulated between the extremes of the EWV and the position that says one should

always stand by one's own belief. But in situations like C.3.1, the spectrum of the EV would disintegrate to only one formulation: the extreme one. Similar reflections hold for all the other views too; therefore, a change from an epistemology of degrees to an epistemology of degrees of belief seems to be appropriate if we want to work with more fine-grained impacts from disagreement on one's doxastic attitude.

The question now is what exactly such an account of degrees of belief is and how we can work with it. So far, we have defined epistemology of degrees only as an approach where there are countless possible doxastic attitudes toward a proposition. More will have to be said in the following section.

3.2 Degrees of Belief and Bayesian Epistemology

Degrees of beliefs are thought to present a level of confidence. Such levels of confidence are also called subjective probabilities. Thus, we can think of a certain degree of belief towards some proposition as the subjective probability that this proposition is true. To formulate this with an example, let us say that Mr. Spock's doxastic attitude concerning some proposition p is 0.74. By definition D.3.2, we can express this using the following equation.

(3.7) $DA_{Spock}(p) = 0.74$

Since we are dealing with subjective probabilities here, we should change this notation. $DA_A(p)$ was meant to be the notation for the doxastic attitude of an agent A towards some proposition p in general. Since in the epistemology of degrees of belief the doxastic attitude is a representation of a probability, we now can use the notation one can find in the literature. This notion for probabilities is $P_A(p)$ and stands for the subjective probability of the agent A towards a proposition p. In this notation, we formulate Mr. Spock's situation with the confidence of 0.74 towards the proposition p with

(3.8) $P_{Spock}(p) = 0.74.$

For a first interpretation, we can say that 3.8 informs us that Mr. Spock considers p to be true with a change of 74%. This function $P_A(p)$, which gives us the doxastic attitude of some agent A towards a proposition p, is called the *probability function*. For a more precise definition of this function, we can use the Kolmogorov (1933) axiomatization.[36]

D.3.3 (Kolmogorov axiomatization): Let Ω be a non-empty set called the *universal set*. The *event space* F is a subset of Ω. F has Ω as a member, and is closed under complementation with respect to Ω and union. Then, a function P from F to the real numbers is a *probability function* if and only if

(1) **Non-negativity:** The probability of an event E element of F is a non-negative real number:

$$P(E) \in \mathbb{R}, P(E) \geq 0 \text{ for all } E \in F$$

(2) **Normalization**: The probability that some elementary event in the entire sample space will occur is 1.

$$P(\Omega) = 1$$

(3) **Finite additivity:**[37] Two events $E_1, E_2 \in F$ with $E_1 \cap E_2 = \emptyset$ satisfy

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

The triple (Ω, F, P) is called the *probability space*.

This definition is nearly what we want. The only problem is that the probability function is not a function with a proposition as an argument. But the probability function P should represent someone's doxastic attitude—i.e. someone's doxastic attitude concerning some proposition p. Therefore, we need a definition of the probability function, where the set F is replaced by a set of propositions. This is not a problem at all,

[36] See Hájek and Hartmann (2010), and Hájek (2012).

[37] Kolmogorov has extended his axiomatization to cover infinite probability spaces and strengthened the third condition to countable additivity: if $E_1, E_2, E_3 \ldots$ is a countably infinite sequence of (pairwise) disjoint sets, each of which is an element of F, then $P(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P(E_n)$.
But since such an extension is controversial, we will not use it here. See Hájek and Hartmann (2010: 4), and Hájek (2012).

since in BE, such a propositional definition of the probability function is the common one.[38]

D.3.4 (Propositional Kolmogorov Axiomatization): Let $\mathcal{A}$ be a set of propositions that is closed under truth-functional combinations and contains all tautologies. The function P from $\mathcal{A}$ to the real numbers is a *probability function* if and only if

(1) The probability that some proposition in the set $\mathcal{A}$ is true is a non-negative real number.

$$P(A) \in \mathbb{R}, P(p) \geq 0 \text{ for all } p \in \mathcal{A}$$

(2) The probability that some tautology T is true is 1.

$$P(\mathrm{T}) = 1 \text{ for any tautology } T \in \mathcal{A}$$

(3) Some incompatible (mutually exclusive) propositions p and q with $p, q \in \mathcal{A}$ satisfy

$$P(p \vee q) = P(p) + P(q)$$

Using these three axioms of the probability function, we also get the following important corollaries. At first, we consider the relation of the probability functions of some proposition p and its negation $\neg p$. The definition of $\mathcal{A}$ ensures us that if $p \in \mathcal{A}$, then it is also the case that $\neg p \in \mathcal{A}$. We also know by propositional logic that a proposition p and its negation $\neg p$ are incompatible, which leads, together with the third axiom, to

(3.9) $P(p) + P(\neg p) = P(p \vee \neg p)$.

But since $p \vee \neg p$ is a tautology, we get by the second axiom that

(3.10) $P(p) + P(\neg p) = 1$.

If we reformulate 3.8, we get the important corollary from D.3.6 that

[38] See Hájek and Hartmann (2010) for the statement that such a propositional definition is the more common one. For the following definition, see Hartmann and Sprenger (2010:3) as well as Hájek (2012).

(3.11) $P(p) = 1 - P(\neg p)$.

Using 3.9, we also get the further corollary that every probability is less than or equal to 1. To prove this, let us reformulate 3.9, which leads us to

(3.12) $P(\neg p) = 1 - P(p)$

Since $\neg p$ is an element of $\mathcal{A}$, we also know from the first axiom in definition D.3.6 that $P(\neg p) \geq 0$ and, therefore, can form the following inequality.

(3.13) $1 - P(p) \geq 0$

$1 \geq P(p)$

But since p is also an element of $\mathcal{A}$, we know by the same axiom that $P(p) \geq 0$ and, therefore, get

(3.14) $0 \leq P(p) \leq 1$ for every proposition $p \in \mathcal{A}$.

BE argues that the doxastic attitude of a rational agent has to satisfy the axioms of D.3.6. This means that the probability function $P(p)$ as defined here describes the *rational degree of belief* towards some proposition p of an agent A. To argue for this, BE uses the so-called *Dutch book argument*, which was introduced by Bruno de Finetti (1927). This argument shows that a violation of the three axioms from D.3.7 would make an agent vulnerable to a so-called Dutch book, a set of bets that guarantees one will lose. This means that an agent whose degrees of belief do not satisfy the axioms above is willing to bet in situations where she cannot possibly win.[39]

It is significant that we are talking about situations where the bettor cannot win through his own subjective probability. The Dutch book argument is not built on the approach that a rational being cannot bet in situations where a Dutch book is given by the objective probabilities. For example, if one's degree of belief in a proposition p is

[39] For a short overview of the Dutch book argument see Easwaran (2013). For proofs, see Kemeny (1955) and Skyrms (1980). Easwaran also provides the outlines for the argument as well as why this "financial problem" is also an "epistemic problem". For example, he presents Skyrms' argument that the Dutch book argument shows that the agent is incoherent. This is the case because the agent values the Dutch book as a sure loss and as something unfavourable, while he also values the bet, which is the Dutch book, as favourable. As Ramsey says, the agents "choice would depend on the precise form in which the options were offered him, which would be absurd" ([1926] 1978: 84).

$P_{subj}(p) = 0.7$, then it is rational for her to accept a bet where she has to pay $0.7€$ on losing the bet and receive $1€$ on winning. This is true even if the real change that p is true is less than 0.7. But it seems odd that the agent would still be considered rational were she to accept such a bet despite knowing this. This is best seen in a situation where $P_{obj}(p) = 0$ because $\neg p$ is a necessary truth. Since $P_{obj}(p) = 0$ and the bettor knows it, she also knows that she will definitely have to pay $0.7€$. If we now assume that she also wants to win, she is as inconsequential in this situation as when she is willing to accept a Dutch book. Hence, Lewis (1980) introduced the so-called *principal principle* (PP).

D.3.5 (Principal Principle): If an agent knows the objective probability of a proposition p to be equal to $P(p)$ and has no "overruling information" available, then the agents' rational degree of belief in p must also be equal to $P(p)$.

Using D.3.6 and D.3.7, we now have a definition of the constraints for the rational degree of belief of an agent towards some proposition p. But the most important part of BE is that it also has the ability to formalize the dynamics of degrees of belief. The tool here is *Bayes' Theorem* (BT). BT was introduced by Thomas Bayes (1701–1761) and is a simple formula for calculating conditional probabilities (cf. Hartman and Sprenger 2010: 1). First, we can introduce the conditional probability (CP) itself. The CP of some proposition p only says how probable it is that this proposition is true, given some body of data E —in other words, how probable it is that both p and E are true, given that E is true. This can be reformulated in formal terms as follows.

D.3.6 (Conditional Probability): The probability that some proposition p is true conditional on a given body of evidence E is

$$P(p|E) = \frac{P(p \wedge E)}{P(E)},$$

provided that both terms of this ratio exist and $P(E) \neq 0$.

Given this definition of CP, we now can deduce BT.

(3.15) $P(E|p) = \frac{P(p \wedge E)}{P(p)}$ $\quad | * P(p)$

$P(p) * P(E|p) = P(p \wedge E)$ $\quad | * \frac{1}{P(E)}$

$\frac{P(p)*P(E|p)}{P(E)} = \frac{P(p \wedge E)}{P(E)}$

But from CF we also know that $P(p|E) = \frac{P(p \wedge E)}{P(E)}$. On inserting this in (3.13) we get

(3.16) $P(p|E) = \frac{P(p)*P(E|p)}{P(E)}$ (Bayes' Theorem)

The idea here is that it is often easier to know the inverse conditional probability $P(E|p)$ than the conditional probability $P(p|E)$ itself. For example, if one wants to know how probable it was for a person that was older than sixty years to die in the year 1984 in Austria, then everything that is necessary to calculate this is available in the form of statistics. Let us say that we call the proposition that some person in Austria died in the year 1984 p_{death}. The information that the person was older than sixty will be represented by E_{60+}. Thus, what we are looking for is the conditional probability $P(p_{death}|E_{60+})$. But by using the death statistics of Austria, one can find out how probable it was that a person that died in this year was older than sixty; hence, we get the inverse conditional probability, $P(E_{60+}|p_{death})$. The statistics of Austria also represent how many persons in Austria were over sixty at the time and how many people died altogether in Austria in that year. Therefore, we also get $P(p_{death})$ and $P(E_{60+})$ from the statistics and have all that we need to calculate $P(p_{death}|E_{60+})$ (cf. Joyce 2008: 2–3).

Using BT, we get the third rationality constraint for an agent's degree of belief, which is formulated by the *Bayesian conditionalization* (cf. Hartmann and Sprenger 2010: 5).

D.3.7 (Bayesian Conditionalization): The rational degree of belief in a proposition p after learning that q is the case, is the conditional probability of p given q:

$$P'(p) = P(p|q)$$

By the means of the Bayesian theorem we get

$$P'(p) = \frac{P(p)P(p|q)}{P(q)}$$

Definitions D.3.6, D.3.7, and D.3.9 bear the three constraints that a rational degree of belief has to satisfy in the light of BE. As Hartman and Sprenger put it:

> Together, the Dutch Book Argument, the Principal Principle and Bayesian Conditionalizations are the three pillars of Bayesian epistemology.
>
> (Harman and Sprenger 2010: 5)

3.3 Disagreement in the Light of Bayesian Epistemology

What was said in the last two sections impacts some of the previously given definitions. In the following, I will change those definitions from chapter 2, which are formulated in terms of belief as a three-fold scheme, and present them in terms of belief as a level of confidence. Let us first consider D.2.1, the definition of disagreement.

D.3.8 (Disagreement II): Two agents A and B disagree about some proposition p if and only if,

(1) $P_A(p) = x$,

(2) $P_B(q) = y$ and B is aware that $P_B(q) = y \rightarrow P_B(p) = z$, where $x \neq z$, and

(4) A and B are aware that they hold differing doxastic attitudes concerning p.

The definition D.2.2 of peerhood still holds because it does not use terms of belief. Since, as per D.2.3, the definition of peer disagreement refers only to D.2.2 and the definition of disagreement, we need not change it either, as long as we now think of disagreement in terms of D.3.3. But there is a further definition that must be changed: the definition of the CV and the NCV.

D.3.9 (CV and NCV): Let us say that two agents A and B are in the situation of peer disagreement over some proposition p.

A view has to satisfy the following conditions in order to be a CV:

(1) $DA'_A(p) = DA'_B(p)$

(2) $DA'_A(p) \neq DA_A(p)$

(3) $DA'_B(p) \neq DA_B(p)$

The NCV is a view that is not a CV.

Condition (1) represents the conformist approach of CV that the agents should overcome the disagreement. Conditions (2) and (3) say that the change has to come from both parties, which means that at least both positions have some weight.

With this definition, we can see that the CV includes the NRD because it implies agreement as defined in the first condition. We also see that this definition does not change the position of the five views in the matrix.

The EWV argues that both doxastic attitudes have the same weight, and one has to take this into account in order to revise one's doxastic attitude. A way to put all this together was proffered by Jehle and Fitelson (2009).[40]

(3.17) $DA'_A(p) = \frac{DA_A(p)+DA_B(p)}{2} = DA'_B(p)$

It is easy to see that a formulation of the EWV as in 3.6 satisfies all three conditions for the CV.

[40] Jehle and Fitelson (2009) argue in this paper that such a formulation of the EWV does not satisfy all the conditions of the view. So, Christensen who has sympathies in Christensen (2007: 203) for something like "splitting the difference", takes in Christensen (2009b), the paper of Jehle and Fitelson (2009) to argue that the middle does not have to be the place where the EWV settles the disagreement, but he does not provide an exact alternative approach. Therefore, I took the approach of Jehle and Fitelson. But does Remark are the reason why I didn't take the condition $DA'_A(p) - DA_A(p) = DA'_B(p) - DA_B(p)$ into the definition D.3.4. Otherwise, if some advocate of the EWV decides that the best result does not have to be the middle between the two doxastic attitudes, the EWV would no longer be a CV, which would be an odd result.

The NSV contradicts the second and third conditions, since both parties can be rational in holding on to their doxastic attitudes as long as their own body of evidence is one that favors this doxastic attitude.

The CRV thinks that one of the opponents of the disagreement has the correct doxastic attitude and should therefore be steadfast. This contradicts—depending on who was correct—the second or the third condition.

The EV and the TEV both allow rational disagreement. Therefore, they are RDV, and since this contradicts the first condition of a CV, they are a NCV.

We can see that the reformulation of the definitions from chapter 2 do not change the relations between the five views that were represented earlier. On the contrary, only such a reformulation allows all of these positions to work in all kinds of cases of peer disagreement because, without it, situations like C.3.1 could not be handled by some of those views since we have a restriction on the possible doxastic attitudes.

4.

Bayesian Epistemology and Defeaters

In the discussion about EoD, the term defeater often plays a role. Peer disagreement is often interpreted as a defeater. For example, Bergmann (2009) tries to distinguish between those situations where peer disagreement is a defeater and those where it is not a defeater. Lackey (2010a and 2010b) argues that the effects of disagreement can be described in terms of a defeater. But if we are talking about disagreement, we have to distinguish between two kinds of defeaters: the *rebutting defeater* and the *undercutting defeater*.[41] The question then is what kind of defeater is a disagreement. Ridder (2013) thinks that disagreements contain both kinds. In addition, Matheson (2009) argues for the influence of disagreement on one's beliefs by drawing parallels to defeaters. Thune (2010), on the other hand, thinks of peer disagreement as a partial defeater, which means that one loses some, but not all of his justification. But not everyone thinks of peer disagreement in terms of a defeater. Another popular position is to describe disagreement, as we did partly above, in terms of HOE, and to distinguish this from a defeater. For example, Christensen (2010) and Weatherson (2013) argue against the similarity between a disagreement and a defeater. All these references make it appropriate to take a closer look at defeaters.

In this chapter, I will begin by presenting the classical definitions of the rebutting defeater and the undercutting defeater by Pollock (1968). The problem is that these definitions are in terms of belief as a three-fold scheme. Nevertheless, I will present cases that intuitively satisfy the conditions for cases of defeaters and then try to analyze those cases in terms of degrees of belief. I hope that my argumentations will match the

[41] For a short overview of this difference, see Kelly (2006).

intuitions of the readers. This method of arguing presents one surplus: it shows the problems of defining the defeater in cases of degrees of belief.

Besides promoting a better understanding of a defeater, this chapter aims to present an analysis of the defeater, which will show that the undercutting defeater and HOE are not contrary in principle, as Christensen (2010) argues. This aim is also my reason for using Pollock's definition, since this is also the definition to which Christensen refers. The thought is to show by a Bayesian analysis of cases of Pollock's defeaters and Christensen's cases of HOE that there is no principal difference.

4.1 Defeater

A defeater is a sort of information that has the power to take away the subject's justification for some belief. Based on Pollock (1986, 37), we can define a defeater as follows.

D.4.1 (Defeater): If R is an epistemic reason for A to believe p, D is a *defeater* for this reason if and only if $(R \wedge D)$ are no longer an epistemic reason for A to believe p.

What Pollock defines here is not a defeater for some agent's doxastic attitude but a defeater for epistemic reasons. In other words, if an agent A believes p on the basis of R, there may be countless information that would satisfy this definition of a defeater, but the agent is still reasonable in holding on to his doxastic attitude. For this to become relevant for the agent, he has to be aware of the information.[42] Therefore, in a situation where an agent knows a defeater, for his epistemic reason he cannot, by definition D.2.7 of epistemic rationality, hold on to his belief. A simple example for such a defeater is the following one.

[42] Or, as in Lackey (2005), even information that one should be aware of, can count as a defeater.

C.4.1 (The Attack): The crew of the Enterprise is locked in a fight. To protect the main window, they have to conceal it with some sort of cover. Therefore, all that they have left to check whether the opponent's ship is destroyed are Mr. Spock's instruments. After some shooting, all the instruments give Mr. Spock reason to believe the proposition $p_{destroyed}$, that the ship is destroyed. Therefore, he forms the corresponding belief $B_{destroyed}$. Since Captain Kirk thinks they are safe again, he gives orders to open the main window. But now they can see with their own eyes that the enemy ship is still in working condition. The information from Mr. Spock's instruments together with the information from the main windows are no longer a reason for Mr. Spock to hold the belief $B_{destroyed}$ since what he sees now is a reason to believe $\neg p_{destroyed}$.

In C.4.1, the information that Mr. Spock gets on seeing the intact ship satisfies the condition you see in D.4.1. This kind of information also satisfies the special type of defeaters that is called the *rebutting defeater* (RD) by Pollock (1986).

D.4.2 (Rebutting Defeater): If R is an epistemic reason for some agent A to believe p, D_r is a *rebutting defeater* for this reason if and only if D_r is a defeater for R and D_r is a reason for A to believe $\neg p$.

Pollock makes a distinction between RD and the *undercutting defeaters* (UD), which are defined as follows (c.f. Pollock 1986, 196).

D.4.3 (Undercutting Defeater): If R is a reason for some agent A to believe p, D_u is an *undercutting defeater* for this reason if and only if D_u is a defeater for R and D_u is a reason for A to doubt or deny that R would not be true unless p were true.[43]

[43] The definitions follow Pollock (1986), but I have changed them slightly. Pollock not only defines D_u as a defeater for R, but as a defeater for believing that R is a reason for A to believe p. If I followed this as well, the following condition, that D_u is a reason for A to doubt that R would not be true unless p were true, would be redundant. However, I think it is analytically clearer to distinguish between these two conditions.

The condition that an agent has to doubt or deny that R would not be true unless p were true in the light of an UD D_u, enunciates that one has to rethink the relation between some information R and a proposition p. To see the difference, we can change C.4.1 in such a way that Mr. Spock has to deal with an undercutting defeater.

C.4.2 (Solar Flares): The crew of the Enterprise is locked in a fight. To protect the main window, they have to conceal it with some sort of cover. Therefore, all that they have left to check whether the opponent's ship is destroyed are Mr. Spock's instruments. After some shooting, all the instruments give Mr. Spock reason to believe the proposition $p_{destroyed}$, that the ship is destroyed. Therefore, he forms the corresponding belief $B_{destroyed}$. At this moment, Lieutenant Commander Montgomery "Scotty" Scott arrives at the bridge and informs Mr. Spock that the ship is currently in an area that has a high number of solar flares, which are causing inaccurate instrumental results. Spock still receives the information from the instruments, but together with the information from Scotty, he no longer has a reason to hold the belief $B_{destroyed}$. But since he also lacks reason to believe that $\neg p_{destroyed}$, he cannot guarantee to Captain Kirk that they are safe again.

The information from Scotty is a UD for Mr. Spock's reason to believe $p_{destroyed}$ because the information from his instruments together with the opinion from the chief engineer Montgomery Scott is no longer a reason to hold this belief. Therefore, the information from Scotty satisfies the condition formulated in D.4.1. But unlike the visual information in C.4.1, Scotty's information is no reason to believe that $\neg p_{destroyed}$t, but only to doubt that the original reason, the information from the instruments, could only be true if the ship were destroyed. Therefore, it also satisfies the second condition formulated in D.4.3.

4.2 Analysis of a Defeater in the Light of Bayesian Epistemology

Cases C.4.1 and C.4.2 seem to be intuitively clear. Nobody would claim that Spock can still believe that the enemy's ship is destroyed even after seeing it intact with his own eyes. It also seems clear that if someone who knows the instruments on the Enterprise as well as the chief engineer Montgomery Scott informed you about negative environmental impacts, you could no longer trust the instruments. But since intuitions do not work anywhere as clearly as here, we have to use such situations to analyze what explains our intuitions and to extract them as parameters for use in cases that lack such clear intuitive answers. In the following, I will use instruments from Bayesian epistemology to do this. Cases C.4.1 and C.4.2 follow the definitions from Pollock (1986) and, therefore, use the three-fold scheme of belief. I will interpret those cases in the formal terms of BE; this includes an approach involving beliefs as an agent's level of confidence. The translation from belief as a three-fold scheme to belief as a level of confidence is a topic that still awaits its solution, but I will start with situations where it seems intuitively clear that the proposed translation is accurate. When this is not the case, I will explicitly refer to the problem of such a translation. This will go hand in hand with the question of how a defeater can be reinterpreted in terms of degrees of belief.

4.2.1 Analyzing Cases Involving the Rebutting Defeater

Example C.4.1 begins with a situation that pushes the intuition that Mr. Spock is justified to believe $p_{destroyed}$. Why is this so? The question is whether it is probable that $p_{destroyed}$ is true simply because the instruments say so. Let us call such a report from Spock's instruments $Rep_{p_{destroyed}}$. We now want to know the probability that just because $Rep_{P_{destroyed}}$ is the case, $p_{destroyed}$ is true. Using the terminology of conditional probability that was discussed in the previous chapter we can reformulate this as

(4.1) $P(p_{destroyed}|Rep_{p_{destroyed}})$,

where (4.1) represents exactly the rational level of confidence that Mr. Spock should have in $p_{destroyed}$ under the condition that $Rep_{p_{destroyed}}$ is the case—i.e. that Mr. Spock's instruments reported that $p_{destroyed}$ is the case. Now, the first step for a solution of 4.1 is to apply the definition D.3.6 of the Bayesian conditionalization so that we get

(4.2) $$\frac{P\left(Rep_{p_{destroyed}}\middle|p_{destroyed}\right)*P(p_{destroyed})}{P(Rep_{p_{destroyed}})}.$$

Let us now take a closer look at 4.2. One of the most important parameters here is $P\left(Rep_{p_{destroyed}}\middle|p_{destroyed}\right)$. This is the probability that Mr. Spock's instruments will report $p_{destroyed}$ if this is the case. Hence, this parameter brings the reliability of the source of the information into the game. That the reliability is important is no surprise. Would we really change C.4.1, and let Mr. Spock flip a coin? We definitely would not think that he has reason to believe that the opponent's ship was destroyed because flipping a coin is not a reliable source for such information. Since the reliability of the source plays such an important role, we define the following parameters.[44]

D.4.4 (True Positive Rate and False Positive Rate): The parameters for the source of the information are:

t the *true positive rate*, which means the probability that we get a report that something is the case if it is the case—i.e. $P(Rep(P)|P)$

f the *false positive rate*, which means the probability that we get a report that something is the case if it is not the case—i.e. $P(Rep(P)|\neg P)$

These two parameters characterize the reliability of the source. We can bring them together in $r := 1 - \frac{f}{t}$. We see that the higher t is, the lower is $\frac{f}{t}$ and, therefore, the higher is r itself. The higher that f is, the higher is $\frac{f}{t}$, and, therefore, the lower is r.

[44] I follow the definition of Bovens and Hartmann (2011: 14). To avoid confusion between the symbols p and q, for the proposition, I changed the names from p for the true positive rate and q for the false positive rate to t and f.

The other parameter is $P(p_{destroyed})$. We will call it the *prior probability* because it is the probability that $p_{destroyed}$ is true independent from the report that $p_{destroyed}$ is true. It is also understandable that this parameter is an important one, since if Spock thinks that the prior probability for $p_{destroyed}$ is 0, maybe because there never was some ship, then too, his instruments could not convince him about $p_{destroyed}$. Hence, Mr. Spock has to admit the possibility of $P_{destroyed}$ prior to his consultation of the instruments. Below, we will talk about the values of such priors.

The denominator of 4.2, $P(Rep_{p_{destroyed}})$, is once again a parameter for the reliability of the source because if we have $P\left(Rep_{P_{destroyed}}\right) = 1$, then the instruments always give the report that $p_{destroyed}$ is correct; in other words, $t = f = 1$. This would be a rather unreliable source. We can now expand this denominator because there are two situations in which we can get the report confirming $p_{destroyed}$: if the ship is destroyed, or if it is not destroyed. Hence, we can have a report saying $p_{destroyed}$, when $p_{destroyed}$ is really the case, or we can have one when it is not the case. Therefore, we get

(4.3) $$P\left(Rep_{p_{destroyed}}\middle|p_{destroyed}\right) * P(p_{destroyed}) +$$

$$P\left(Rep_{p_{destroyed}}\middle|\neg p_{destroyed}\right) * P(\neg p_{destroyed}).$$

Since we have already defined parts of this sum in D.4.3, we can reformulate 4.3 as

(4.4) $$t * P(p_{destroyed}) + f * P(\neg p_{destroyed}).$$

If we now reinsert the denominator in 4.2 and apply D4.3, we get

(4.5) $$\frac{t*P(p_{destroyed})}{t*P(p_{destroyed})+f*P(\neg p_{destroyed})} = \frac{t*P(p_{destroyed})}{t*P(p_{destroyed})+f*\left(\frac{t}{t}\right)*P(\neg P_{destroyed})} =$$

$$\frac{t*P(p_{destroyed})}{t*P(p_{destroyed})+t*\left(\frac{f}{t}\right)*P(\neg p_{destroyed})} = \frac{t*P(p_{destroyed})}{t*\left(P(p_{destroyed})+\left(\frac{f}{t}\right)*P(\neg p_{destroyed})\right)} =$$

$$\frac{P(p_{destroyed})}{P(p_{destroyed})+\frac{f}{t}*P(\neg p_{destroyed})}$$

Now, let us try to represent C.4.1 by applying the formula 4.5. We must first think about some reasonable values for the parameters. Let us start with the reliability of Spock's instruments. We know that the instruments on the Enterprise are highly sophisticated. Thus it is reasonable to think that Mr. Spock knows that his instruments nearly always show him p if it is the case—say, 90% of the time; therefore, we get $t_{instruments} = 0.90$.[45] But better than the true positive rate of the instruments on the Enterprise is the false positive rate, which means that Mr. Spock knows that his instruments nearly never show him p if p is not the case; therefore, we set the false positive rate as $f_{instruments} = 0.05$. Now, we just need the prior probability for $p_{destroyed}$. Since Spock knows nothing about the ship, especially not how well the weapons of the Enterprise will work on it, we can say that he is totally unsure about the status of the opponent's ship. Therefore, we set $P^0(p_{destroyed}) = 0.5$.[46] The prior probability that the ship is not destroyed is therefore by definition D.3.4 $P^0(p_{\neg destroyed}) = 1 - P^0(p_{destroyed}) = 0.5$. These values give us the following result for 4.5.

(4.6) $$P^1(p_{destroyed}) = \frac{P^0(p_{destroyed})}{P^0(p_{destroyed}) + \left(\frac{f_{instruments}}{t_{instruments}}\right) * P^0(\neg p_{destroyed})} = \frac{0.5}{0.5 + \frac{0.05}{0.90} * 0.5}$$

$$P^1(p_{destroyed}) = 0.9474$$

Thus, we can say that Spock is quite sure that the opponent's ship is destroyed, after he consults his highly reliable instruments. But what happens when he sees with his own eyes that the ship is not destroyed? The eyes are now his instruments, and as is commonly known, if there is no bigger problem like darkness or a dust grain, we consider visual perception to be highly trustworthy too. Mr. Spock trusts his eyes so much that he

[45] As mentioned above, the advantage of the Bayesian approach of conditional probabilities is that we can reduce it to probabilities like the value of t and f. These need not be only subjective probabilities but can also be the result of statistical records.

[46] This is only one way to argue for the values of the priors. There are two positions used by advocates of BE regarding the priors. One thinks that since priors have no conditions at all, every value is a reasonable one. This position is called subjective Bayesianism. On the other side, we have objective Bayesianism, which holds that priors are determined by symmetry constraints as were presented in an argument for $P^0(p_{destroyed}) = 0.5$. In this work, I will follow the constraints of objective Bayesianism. One of the most famous examples of an argumentation for such an objective Bayesianism can be found in Jaynes (1968). For an overview of this debate, see Talbott (2011).

would think they nearly never show him something that is not there and nearly always if it is. So it seems reasonable to set $t_{eyes} = 0.99$ and $f_{eyes} = 0.01$.[47] The prior probability is now $P^1(p_{destroyed})$, because Mr. Spock has now already consulted his instruments and become very sure that the ship is destroyed. But we are still left with a problem. We are not interested in the post probability after a positive report, but after a negative one. Therefore, we are looking for

(4.7) $P\left(P_{destroyed} | Rep_{\neg P_{destroyed}}\right)$.

Here, we can apply the Bayesian conditionalization again and get

(4.8) $\frac{P\left(Rep_{\neg P_{destroyed}} \middle| P_{destroyed}\right) * P(P_{destroyed})}{P(Rep_{\neg P_{destroyed}})}$.

To move on from here, we have to take another closer look at the parameters. At first, $P\left(Rep_{\neg p_{destroyed}} \middle| p_{destroyed}\right)$. If we make the working definition that $q \coloneqq \neg p_{destroyed}$, and apply that p has the same truth value as $\neg\neg p$, we come to $P\left(Rep_q \middle| \neg q\right)$. From D.4.3, we know that this is the probability value of f. If we now apply the working definition on the denominator, we get $P(Rep_q)$.[48] We can reform it analogous to 4.3, apply D.4.3, and get

(4.9) $P\left(P(Rep_q) \middle| \neg q\right) * P(\neg q) + P\left(P(Rep_q) \middle| q\right) * P(q) =$

$f * P(\neg q) + t * P(q)$

If we now reinsert $\neg p_{destroyed}$ for q and place all this in 4.8, we get

(4.10) $\frac{f * P(p_{destroyed})}{f * P(\neg p_{destroyed}) + t * P(\neg p_{destroyed})} = \frac{f * P(p_{destroyed})}{f * P(p_{destroyed}) + t * \left(\frac{f}{f}\right) * P(\neg p_{destroyed})} =$

[47] Again, those values can be just Mr. Spock's subjective probabilities towards the reliability of his eyes, but if he were to know about objective probabilities from studies, he would have to take these values as his subjective probabilities, if he wanted to be rational. This is what the principal principle from definition D.3.5 says.

[48] This is a simplification because the reliability for the instrument to detect the negation of some proposition p does not have to be the same as the reliability towards the proposition itself. Besides the true positive rate and the false positive rate, we also have the true negative rate and the false negative rate. To keep things simple and since it does not change the basic ideas behind what will be said, we can ignore this.

$$\frac{f*P(p_{destroyed})}{f*P(p_{destroyed})+f*\left(\frac{t}{f}\right)*P(\neg p_{destroyed})} = \frac{f*P(P_{destroyed})}{f*\left(P(p_{destroyed})+\left(\frac{t}{f}\right)*P(\neg p_{destroyed})\right)} =$$

$$\frac{P(p_{destroyed})}{P(p_{destroyed})+\frac{t}{f}*P(\neg p_{destroyed})}.$$

We can now see that the difference between 4.5 and 4.10 is the inverse value of the reliability parameters. If we apply this formula on our values, we get the following result.

(4.11) $$P^2(p_{destroyed}) = P^2\left(P_{destroyed}|Rep_{\neg p_{destroyed}}\right) =$$

$$= \frac{P^1(p_{destroyed})}{P^1(p_{destroyed})+\left(\frac{t_{eyes}}{f_{eyes}}\right)*P^1(\neg p_{destroyed})}$$

$$P^2(p_{destroyed}) = 0.1539$$

We can now see that the formal analysis of the case C.4.1 confirms our intuition. Mr. Spock is initially rational in having high confidence in $p_{destroyed}$, as we would assume. But after seeing the opponent's ship and, therefore, receiving the report confirming $Rep_{\neg p_{destroyed}}$, he is compelled to decrease his confidence. Let us now take a closer look at the second case, C.4.2.

4.2.2 Analyzing Cases Involving the Undercutting Defeater

The beginning of C.4.2 is the same as in C.4.1; therefore, we once again work with the values $t_{instruments} = 0.90$, $f_{instruments} = 0.05$, and $P^0(p_{destroyed}) = 0.5$. Since we can also apply formula 4.5, we start again with the post probability $P^1(p_{destroyed}) = 0.9474$. From here on, C.4.2 differs from C.4.1 because now there is no information about $p_{destroyed}$, but there is information about the reliability values $t_{instruments}$ and $f_{instruments}$. Hence, the next step is to analyze how those values change after Scotty's report. Before that, we have to think about what Scotty is reporting.

4.2.2.1 Reports about Reliability

One way to work with Scotty's report is by following Bovens and Hartmann (2011: 56–88), and introducing a single variable relating to the reliability of a source.

D.4.5 (Reliability): The dual variable Rel_S is true if the source S is reliable and false if S is not reliable, where reliable means that if something is the case, the source S would give a report about it, and if not it would not. Therefore, we get:

$$P(Rep_p|p, Rel_S) = 1 \text{ and } P(Rep_p|\neg p, Rel_S) = 0$$

We also say that if a source is not reliable, it is as likely to report that p is true in those situations where p is the case as in those where p is not the case. Therefore, we get:

$$P(Rep_p|p, \neg Rel_S) = P(Rep_p|\neg p, \neg Rel_S) =: a$$

If we now consider Scotty's report as a report that the instruments are not reliable, then he is saying that $\neg Rel_{instruments}$, is true. Hence, we get the report $Rep_{\neg Rel_{instruments}}$. The question now is how sure we can be that the instruments are reliable following this report. This means that we have to determine the following conditional probability.

(4.12) $P(Rel_{instruments}|Rep_{\neg Rel_{instruments}})$

The situation here is now the same as in 4.7. Therefore, we can also apply formula 4.10 and get

(4.13) $P(Rel_{instruments}|Rep_{\neg Rel_{instruments}}) = \frac{P(Rel_{instruments})}{P(Rel_{instruments})+\left(\frac{t}{f}\right)*P(\neg Rel_{instruments})}$.

In the following, we have to consider the values of the parameters of this formula.

4.2.2.2 Reliability, True Positive Rate, and False Positive Rate

Until now, all that we know about what Spock is thinking regarding his instruments are t and f, the true positive rate and the false positive rate. But we have to know how sure he is that his instruments are reliable, because $P(Rel_{instruments})$ is the prior probability that Spock holds regarding the reliability of his instruments. The question therefore is whether there is a relationship between those parameters. The answer is yes (cf. Bovens and Hartman 2003, p. 74). Let us start with t and its relation to $P(Rel_{instruments})$:

(4.14) $t = P(Rep(p)|p)$ by D.4.3

$= \sum_{Rel} P(Rep(p), Rel|p)$ by expansion

$= \sum_{Rel} P(Rep(p)|Rel, p)P(Rel|p)$ by the chain rule

$= \sum_{Rel} P(Rep(p)|Rel, p)P(Rel)$ by independence[49]

$= P(Rep(p)|Rel, p)P(Rel) + P(Rep(p)|\neg Rel, p)P(\neg Rel)$

$= 1 * P(Rel) + a * P(\neg Rel)$ by D.4.5

$t = P(Rel) + a * (1 - P(Rel))$

Now, we can do the same for f:

(4.15) $f = P(Rep(p)|\neg p)$ by D.4.3

$= \sum_{Rel} P(Rep(p), Rel|\neg p)$ by expansion

$= \sum_{Rel} P(Rep(p)|Rel, \neg p)P(Rel|\neg p)$ by the chain rule

$= \sum_{Rel} P(Rep(p)|Rel, \neg p)P(Rel)$ by IA

$= P(Rep(p)|Rel, \neg p)P(Rel) + P(Rep(p)|\neg Rel, \neg p)P(\neg Rel)$

$= 0 * P(Rel) + a * P(\neg Rel)$ by D.4.5

[49] I borrowed this condition from Boven and Hartmann (2003:56-80), where they introduced it as an innocent assumption since it only says that knowing about the reliability of a source alone says nothing about the truth of the proposition; only the report counts. I will refer to this assumption as IA.

$$f = a * (1 - P(Rel))$$

If we now apply 4.15 on 4.14, we get

(4.16) $t = P(Rel) + f$

$$t - f = P(Rel)$$

$$P(Rel) = t - f$$

4.2.2.3 Reliability Reports and Their Effects

Since we know that $t_{instruments} = 0.90$ and $f_{instruments} = 0.05$, using 4.16 we now get $P(Rel_{instruments}) = 0.85$. To work with 4.13, we now have to think about the reliability of Scotty. Mr. Spock knows as well as we do that Scotty is a good chief engineer. Therefore, he believes that Scotty knows what he is talking about. Let us say that Mr. Spock would agree with the following attribution: $t_{Scotty} = 0.85$ and $f_{Scotty} = 0.1$. Since we know that $P(\neg Rel_{instruments}) = 1 - P(Rel_{instruments})$, we can apply those values on 4.13.

(4.17) $P(Rel_{instruments}|Rep_{\neg Rel_{instruments}}) = \frac{P(Rel_{instruments})}{P(Rel_{instruments}) + \left(\frac{t_{Scotty}}{f_{Scotty}}\right) * P(\neg Rel_{instruments})} =$

$$\frac{P(Rel_{instruments})}{P(Rel_{instruments}) + \left(\frac{t_{Scotty}}{f_{Scotty}}\right) * (1 - P(Rep_{instruments}))} = \frac{0.85}{0.85 + \left(\frac{0.85}{0.1}\right) * (1 - 0.85)} =$$

$$P(Rel_{instruments}|Rep_{\neg Rel_{instruments}}) = 0.4$$

This result follows our intuition because after a report about the negative environmental impacts on the instruments from the chief engineer, Mr. Spock would be ignorant not to lower his confidence that the instruments are reliable. We must now translate this value back into the parameters of true positive rate and false positive rate. 4.15 already tells us that $f = a * (1 - P(Rel))$, but to apply this formula we have to determine the value for a, . Since we know that the prior value for $P(Rel_{instruments}) = 0.85$ and that the prior value of $f_{instruments} = 0.05$, we get the following calculation for a:

(4.18) $f_{instruments} = a * (1 - P(Rel_{instruments}))$

$$\frac{f_{instruments}}{(1-P(Rel_{instruments}))} = a$$

$$a = \frac{f_{instruments}}{(1-P(Rel_{instruments}))} = \frac{0.05}{1-0.85} = 0.\dot{3}$$

Since in this interpretation Scotty only comments on the reliability and not the probability that the instruments might give a positive report at all, this value does not change after Scotty's report. Thus, we can now apply this value, together with the post value $P^1(Rel_{instruments})$ on 4.15 and 4.16, and get the post values for ${t^1}_{instruments}$ and ${f^1}_{instruments}$.

(4.19) ${t^1}_{instruments} = P^1(Rel_{instruments}) + a * (1 - P^1(Rel_{instruments}))$

$${t^1}_{instruments} = 0.4 + 0.\dot{3} * (1 - 0.4)$$

$${t^1}_{instruments} = 0.6$$

(4.20) ${f^1}_{instruments} = a * (1 - P^1(Rel_{instruments}))$

$${f^1}_{instruments} = 0.\dot{3} * (1 - 0.4)$$

$${f^1}_{instruments} = 0.2$$

To see how this influences Spock's confidence that the opponent's ship was destroyed, we have to apply these new values on 4.5 and get the following result.

(4.21) $P^2(p_{destroyed}) = \frac{P^0(p_{destroyed})}{P^0(p_{destroyed}) + \frac{{f^1}_{instruments}}{{t^1}_{instruments}} * P^0(\neg p_{destroyed})} = \frac{0.5}{0.5 + \frac{0.1714}{0.6571} * (1-0.5)} =$

$$P^2(p_{destroyed}) = 0.75$$

After 4.11, we said that the results match our intuition about the rebutting defeater. But 4.21 seems to not do the same for our intuitions about the undercutting defeater. Even after Mr. Spock gets information from Scotty that the instruments are no longer reliable, he reduces his trust in them, but the result is still a confidence of 75% that the opponent's ship is destroyed. If we take a look at D.4.3, the definition of an undercutting

defeater, we see two conditions. One is that an undercutting defeater is a reason for an agent to doubt or deny that the reason on which he built his belief would not be true unless what he believes were true. We can see that our analysis fulfils this condition. We said that Mr. Spock reduces his confidence in the instruments. This means that he no longer thinks that what the instruments indicate have such a strong connection to the actual case. By reducing his trust in the instruments, he reduces his trust in them as a reason to believe that the opponent's ship is destroyed. Therefore, he sees Scotty's announcement as a reason to doubt what the instruments would not show him what they do unless the ship really was destroyed. So the information that Mr. Spock receives from Scotty fulfils this condition to be an undercutting defeater. But what about the second one, the condition to be a defeater, which is defined by D.4.1? Following D.4.1, Scotty's announcement only counts as a defeater if what Mr. Spock gets from his instruments together with the information from Scotty no longer count as a reason to believe that the opponent's ship is destroyed.

After presenting case C.4.2., I argued that intuition suggests that Scotty's report should count as a defeater. But our analysis does not seem to capture this intuition. In the following, I will give two reasons why this is the case. First, I will talk about a general problem that has to do with D.41 and the conception of beliefs as probability values. But this alone will not fix the gap between intuition and the analysis concerning case C.4.2. Therefore, I will then take a closer look at Scotty's report and present an alternative interpretation.

The first point arises because Pollock (1986) lacks a probabilistic conception of beliefs. Therefore, we have to think about the influence of the change to a probabilistic conception on his definition of a defeater. In D.4.1, the condition to be a defeater was that the reason for believing a proposition p, together with the defeater, are no longer a reason to believe p. Since belief is for Pollock's conception as defined in D.3.2, we only have three possible doxastic attitudes. There is the positive one (i.e. the belief that some proposition p is the case), the negative one (i.e. the belief that some proposition p is not the case), and the agnostic one (i.e. one neather believes that p the case nor that it is not). Moreover, since the defeater has to take away the reason to hold on to the positive

state, the minimum condition for a defeater is the change from believing to being agnostic towards some proposition p, as this is the minimum change in this conception.

One way to translate this into a view with a probabilistic conception is to start with an interpretation of the agnostic state of believing. One could say that agnosticism can be captured only by being totally unsure about the proposition, which means that the probability that the proposition is true is the same as the probability that the proposition is wrong. Since the probability for the negation of a proposition p is $1 - P(p)$, the only value where this is the case is 0.5. So, to be agnostic about a proposition means to give it a probability of 0.5. With such a probabilistic translation for agnosticism, one can now argue that since a defeater is defined to change the belief status at least to agnosticism, a defeater in probabilistic terms has to reduce confidence in a proposition to 0.5 or less. If this were true, the results from 4.21 would show that Scotty's report is definitely not a defeater, and, therefore, also not an undercutting defeater.

This would be one way to think about a defeater in a probabilistic context. The other way would be one, where we start to interpret a defeater not by a translation of agnosticism, but by D.4.1 itself. This definition says nothing about agnosticism, but only that a defeater takes away one's reason to believe some proposition. But since we are no longer using a three-valued conception of belief, one must ask what this means in probabilistic terms. Is the only correct interpretation really the reduction to exactly 0.5 or would a reduction to around 0.5 also do the job?

To say that believing is a certainty of more than 0.5, not believing a certainty of less than 0.5, and agnosticism a certainty of exactly 0.5 cannot be the right translation, because then one has to question the need to work with a probabilistic conception of belief at all.[50] Therefore, the better interpretation of D.4.1 in probabilistic terms seems to be one, where a defeater has to make a drastic reduction of one's probabilistic state of believing. The question for the case above will now be whether the reduction of Mr. Spock's certainty in the proposition that the opponent's ship is destroyed is drastic

[50] See Foley (1992), where he is discussing the problems of a translation from the three-fold scheme of belief to degrees of belief.

enough. The problem of answering this question is the problem that always accompanies such vague terms—namely, that there cannot be just one correct answer.

On the on hand, it is a reduction from a very high level of nearly 0.95 to 0.75. It seems clear that Scotty is a reliable source, but one cannot consider his announcements to be totally certain; Mr. Spock does not have to drop all his trust in his instruments. He has to reduce his trust, but not lose it all. Therefore, his instruments will still raise his confidence in the proposition that the opponent's ship is destroyed. The confidence will not be as high as it would have been without Scotty's report, but it will be above Mr. Spock's prior certainty of 0.5. Why can there still be some strong intuition that this reduction is not enough? After all, Mr. Spock also has high confidence that Scotty is telling the truth. And the latter is saying that Mr. Spock's instruments are no longer reliable.

Earlier, I presented an interpretation for Scotty's announcement: that Scotty is essentially telling Mr. Spock that his instruments are as certain as flipping a coin. But is this really always what we want to say, when we say that something is not reliable? If we know a person who nearly always lies about a given topic, would we not describe the person as unreliable in this field in order to warn somebody? By this, we would not want to say that the person is as reliable as flipping a coin, but worse. Not only is she *not-reliable*, but she is also *anti-reliable*—i.e. it is more certain that what she is saying is untrue, than true.[51] It seems to me that when the chief engineer arrives at the bridge and says that the instruments do not work in this environment, it is legitimate to think that he is telling us that the instruments are not as reliable as flipping a coin, but worse. Therefore, we can start the interpretation of C.4.1 from this point of view again.

4.2.2.4 Reports about Anti-reliability

D.4.5 defines the variable Rel, which if true, says about an agent A that this agent will say what is the case, and not say what is not the case. D.4.5 also says that if somebody is not a reliable source, then $\neg Rel$ is true for that agent; in that case, the reports of this agent

[51] A similar distinction was made by Egan and Elga (2005).

are as likely to be true as wrong. Above, Scotty's report about the impact on the instruments was interpreted as unreliable in this sense. Given this new point of view, Scotty wanted to say that the instruments are more likely to tell Mr. Spock something wrong, and so they have to think about a new way to fit this announcement into the probabilistic analysis.

Scotty is saying that the instruments will say about what is not the case, that it is, and about what is the case, that it is not. Therefore, we can split his report into two: one part representing the true positive rate, and the other the false negative rate (see D.4.3). Therefore, we can define the following two variables:

D.4.6 (Anti-Reliability): The dual variable Rel_S^+ is true if the source S tells always about what is the case, that it is the case. Therefore we get:

$$P\left(Rep_p \middle| p, Rel_S^+\right) = 1$$

The dual variable Rel_S^- is true if the source S tells never about what is not the case, that it is the case. Therefore, we get:

$$P\left(Rep_p \middle| \neg p, Rel_S^-\right) = 0$$

We also define the negations of these two variables as follows:

$$P\left(Rep_p \middle| p, \neg Rel_S^+\right) = 0$$

$$P\left(Rep_p \middle| \neg p, \neg Rel_S^-\right) = 1$$

If for some source S, $\neg Rel_S^+$ holds as well as $\neg Rel_S^-$, we call this source anti-reliable.

Using this interpretation, it appears that Scotty is saying that $\neg Rel_{instruments}^+$ and $\neg Rel_{instruments}^-$ are true. Hence, we get from Scotty the two reports $Rep_{\neg Rel_{instruments}^+}$ and $Rep_{\neg Rel_{instruments}^-}$. The question now is how sure Mr. Spock can be regarding whether the instruments are reliable after this report. This means that we have to determine the following conditional probabilities:

4.22 $$P\left(Rel^{+}_{instruments}|Rep_{\neg Rel^{+}_{instruments}}\right) = \frac{P(Rel^{+}_{instruments})}{P(Rel^{+}_{instruments})+\left(\frac{t}{f}\right)*P(\neg Rel^{+}_{instruments})}$$

4.23 $$P\left(Rel^{-}_{instruments}|Rep_{\neg Rel^{-}_{instruments}}\right) = \frac{P(Rel^{-}_{instruments})}{P(Rel^{-}_{instruments})+\left(\frac{t}{f}\right)*P(\neg Rel^{-}_{instruments})}$$

Now, we have to think about the values of those parameters.

4.2.2.5 Anti-reliability, True Positive Rate and False Positive Rate

In 4.2.2.2, I was searching for a connection between the true positive rate, the false negative rate and $P(Rel_{instruments})$. Here, I will do the same, but between the true positive rate, the false negative rate, and the $P(Rel^{+}_{instruments})$ and the $P(Rel^{-}_{instruments})$ respectively. As above, we start with what we know about t, the true positive rate:

(4.24) $t = P(Rep(p)|p)$ by D.4.3

$= \sum_{Rel} P(Rep(p), Rel^{+}|p)$ by expansion

$= \sum_{Rel} P(Rep(p)|Rel^{+}, p)P(Rel^{+}|p)$ by the chain rule

$= \sum_{Rel} P(Rep(p)|Rel^{+}, p)P(Rel^{+})$ by IA

$= P(Rep(p)|Rel^{+}, p)P(Rel^{+}) + P(Rep(p)|\neg Rel^{+}, p)P(\neg Rel^{+})$

$= 1 * P(Rel^{+}) + 0 * P(\neg Rel^{+})$ by D.4.6

$t = P(Rel^{+})$

Now, we can do the same for q, the false negative rate:

(4.25) $f = P(Rep(p)|\neg p)$ by D.4.3

$= \sum_{Rel} P(Rep(p), Rel^{-}|\neg p)$ by expansion

$= \sum_{Rel} P(Rep(p)|Rel^{-}, \neg p)P(Rel^{-}|\neg p)$ by the chain rule

$$= \sum_{Rel} P(Rep(p)|Rel^-, \neg p)P(Rel^-) \quad \text{by IA}$$

$$= P(Rep(p)|Rel^-, \neg p)P(Rel^-) + P(Rep(p)|\neg Rel^-, \neg p)P(\neg Rel^-)$$

$$= 0 * P(Rel^-) + 1 * P(\neg Rel^-) \quad \text{by D.4.6}$$

$$f = P(\neg Rel^-)$$

4.2.2.6 Anti-reliability Reports and Their Effects

Since we are still working on the same case, the values remain the same as well. Therefore, we already know that $t_{instruments} = 0.90$ and $f_{instruments} = 0.05$. By 4.24 and 4.25, we now also get $P(Rel^+_{instruments}) = 0.90$, $P(\neg Rel^+_{instruments}) = 1 - P(Rel^+_{instruments}) = 0.1$, $P(\neg Rel^-_{instruments}) = 0.05$, and $P(Rel^-_{instruments}) = 1 - P(\neg Rel^-_{instruments}) = 0.95$. The reliability values for Scotty are also still the same; thus, we have $t_{Scotty} = 0.85$ and $f_{Scotty} = 0.1$. We can now apply these values to 4.22 and 4.23.

(4.26) $$P\left(Rel^+_{instruments}|Rep_{\neg Rel^+_{instruments}}\right) = \frac{P(Rel^+_{instruments})}{P(Rel^+_{instruments}) + \left(\frac{t_{Scotty}}{f_{Scotty}}\right) * P(\neg Rel^+_{instruments})}$$

$$= \frac{0.90}{0.90 + \left(\frac{0.85}{0.1}\right) * 0.10}$$

$$P\left(Rel^+_{instruments}|Rep_{\neg Rel^+_{instruments}}\right) = 0.5143$$

(4.27) $$P\left(Rel^-_{instruments}|Rep_{\neg Rel^-_{instruments}}\right) = \frac{P(Rel^-_{instruments})}{P(Rel^-_{instruments}) + \left(\frac{r_{Scotty}}{f_{Scotty}}\right) * P(\neg Rel^-_{instruments})}$$

$$= \frac{0.95}{0.95 + \left(\frac{0.85}{0.1}\right) * 0.05}$$

$$P\left(Rel^-_{instruments}|Rep_{\neg Rel^-_{instruments}}\right) = 0.69\dot{0}\dot{9}$$

With 4.26 and 4.27, and together with 4.24 and 4.25, we now get the post values for the true positive rate and the false negative rate after Scotty's report.

(4.28) $$t^1{}_{instruments} = P^1(Rel^+_{instruments}) = 0.5143$$

(4.29) $f^1{}_{instruments} = 1 - P^1(Rel^-_{instruments}) = 0.30\dot{9}$

To see how this influences Mr. Spock's confidence that the opponent's ship was destroyed, we have to apply those new values on 4.5 and get the following result.

(4.30) $P^2(P_{destroyed}) = \frac{P(P^0{}_{destroyed})}{P(P^0{}_{destroyed}) + \frac{f^1{}_{instruments}}{t^1{}_{instruments}} * P(\neg P^0{}_{destroyed})} = \frac{0.5}{0.5 + \frac{0.30\dot{9}}{0.5143} * (1 - 0.5)} =$

$P^2(P_{destroyed}) = 0.6246$

We can see in 4.30 that the report of the anti-reliability of the source has much more impact than merely the report from its non-reliability. This result should not surprise us, but the question still remains whether our analyses go hand in hand with our intuition. In C.4.2, we said that Mr. Spock no longer has a reason to hold the belief $B_{destroeyed}$, but neither does he have a reason to believe the opposite. This agnostic status exists in a system where there are only the three modes of believing, agnosticism and the believing of the opposite of the minimal change that one can require from a defeater. If we think that agnosticism is also the minimal change, even if we do not think in the frame of such three-fold belief systems, and that agnosticism is only the case if one's certainty that p is the case is the same as the certainty that p is not the case, it means $P(p) = 0.5$, ; then too, the report of anti-reliability does not count as a defeater. On the other hand, as I mentioned above, one may be better off thinking of a defeater as some drastic reduction in one's confidence that something is the case. Such a drastic reduction occurs in the situation where Mr. Spock gets the anti-reliability report on the case, because he was nearly absolutely certain that $p_{destroyed}$ was true; now, he only has the confidence $P^2(p_{destroyed}) = 0.6246$. Granted, he still has a little more confidence in $p_{destroyed}$ than in $\neg p_{destroyed}$, but after the change he is definitely nearer to 0.5, whereas before the change he was almost absolutely certain. Therefore, it seems appropriate in the light of 4.30 to say that Mr. Spock is defeated.[52]

[52] In the summary, I will say a little more about my opinions regarding the relation between a defeater and the concept of beliefs as degrees of confidence. In this case, I hope one will accept my motivation for Mr. Spock's being defeated after the anti-reliability report.

5

Bayesian Epistemology and HOE

As mentioned at the beginning of the chapter above, Christensen (2010) thinks of HOE as something that is principally different from UD. He argues that HOE leads to a situation where "I must in some sense, and to at least some extent, *put aside* or *bracket* my original reasons" (Christensen 2010, 195). He further argues that one has to do so "although I have conclusive evidence for the correctness" (196) of the proposition in which I believe. I think he is wrong here. I think that HOE attacks one's evidence by lowering the probability that the evidence is correctly evaluated. The difference is important, because Christensen's way of thinking about HOE bears one of the major distinctions between the EWV and the TEV. While the EWV only takes the equality of the situation of disagreement into account, the TEV also argues for taking one's FOE that leads to the disagreement into account. That the EWV is doing so seems to be the consequence of an account of HOE as something that *put aside* my evidence. Therefore, in arguing against this position I am also arguing for a TEV because I believe that one's original FOE will have to be re-evaluated in the situation of peer disagreement, because of the HOE that the disagreement is, but not *bracket*.

To argue for this, I started above with a Bayesian analysis of a defeater. I will now go on and do the same for HOE. The aim of this chapter together with the previous chapter is to show that there is no principal difference between UD and HOE. Since Christensen is arguing for this difference by arguing for the putting-aside force of HOE, I will reject this force by arguing against the principal difference. I will present the cases of HOE that Christensen (2010) uses and do the same as above with the cases of defeater. In the next

chapter, I will use the same analyses once again to show that if Christensen takes his interpretation of peer disagreement as HOE seriously he has to give up his EWV.

5.1 Higher Order Evidence

In chapter 2, we had already heard about HOE. Christensen argues that peer disagreement is HOE, but he also presents further cases to illustrate what HOE is and what it does.[53]

C.5.1 (Sleep Deprivation): Dr. Leonard "Bones" McCoy is the chief medical officer aboard the USS Enterprise. After diagnosing a particular patient's condition and prescribing certain medications, he is reminded by a nurse that he has been awake for 36 hours. Knowing by some highly reliable study that one is prone to making cognitive errors in such circumstances, he reduces his confidence in his diagnosis.

Christensen argues along with Feldman (2005) that cases like C.5.1 are different from cases of the undercutting defeater—i.e. cases like C.4.2. Their line of argumentation is that the undercutting defeater shows only that the actual reasoning cannot be relied on in the present case, while HOE denies that there is an evidential connection at all (c.f. Feldman 2005, p. 113 and Christensen 2010, p. 194). I argue that the undercutting defeater changes the connection between the evidence and the degree of one's confidence by changing the reliability of that evidence. And yes, reliability is always a parameter that only holds for some time and place; therefore, by changing the circumstances, one changes the reliability. If Feldman and Christensen argue about the undercutting defeater, that there is no connection only in the present case, I have to agree. This, however, is just to say that the reliability of the source drops in the current environment. But in the case of HOE, we also have to re-evaluate the reliability. This may not happen because of some environmental change, but by changes in the source of my evidence itself. Like in C.5.1, Bones' medical evaluations are normally very reliable.

[53] I took the cases from Christensen and only changed the characters, so that we do not have to leave the Star Trek universe (c.f. Christensen 2010, p. 186 sq.).

But since the instrument of those evaluations, Bones himself, is in bad shape, the reliability is changing. This does not mean that one has to deny the connection between Bones evaluations more than one has to deny such a connection between Mr. Spock's instruments and the evaluation of the circumstances in C.4.2. While it is true that the connection is not as strong as before, this is exactly what the reliability change is telling us. By reducing the reliability of some source, one is reducing his trust that the source and the truth about a particular topic have some sort of connection. If one is reducing the reliability ascribed to the source because of a change in the environment or because of a change in the source is unrelated to the consequences. So, perhaps we have to distinguish between two kinds of undercutting defeaters—the *environmental undercutting defeater*, and the *source undercutting defeater*. We could also call those defeaters as HOE, and distinguish between *HOE about the environment* and *HOE about the source*. But just because we have some different names, it does not mean that we are talking about different things. Let us analyze C.5.1 to show that that there is some analogy with C.4.2.

5.2 Analyzing Higher Order Evidence

Example C.5.1 again begins with a situation where some person—in this case, Bones—seems to be justified in his belief. This is the case because he is the chief medical officer, and we could argue that he would have to be a reliable source on medical topics. Let us say that Bones knows that he is such a reliable source and would rate his reliability with $t_{Bones} = 0.8$ and $f_{Bones} = 0.2$. Before examining a patient he has no idea what the diagnosis will be; therefore, he has to ascribe his diagnosis—let us call it diagnosis X—to the prior probability $P(p_{diagnosis\,X}) = 0.5$. Now, we can evaluate the confidence that Bones should have in his diagnosis, by applying 4.5, and getting the following result.

(5.1) $$\frac{P(P_{diagnosis\,X})}{P(P_{diagnosis\,X})+\frac{q_{Bones}}{p_{Bones}}*P(\neg P_{diagnosis\,X})} = \frac{0.5}{0.5+\frac{0.2}{0.8}*0.5} = 0.\dot{8}$$

But now, Bones has the information from some study that people who, like him, have been awake for 36 hours are no longer reliable. As this study is very reliable, let us say that $t_{study} = 0.85$ and $f_{study} = 0.15$. To keep things simple, let us say the study shows

through these reliability values that Bones is anti-reliable. This means that we have to revaluate his reliability analogous to 4.26–4.29. Thereby, we get Bones' post-reliability values.

(5.2) $$t^1_{Bones} = P\left(Rel^+_{Bones} | Rep_{\neg Rel^+_{Bones}}\right) = \frac{P(Rel^+_{Bones})}{P(Rel^+_{Bones}) + \left(\frac{t_{study}}{f_{study}}\right) * P(\neg Rel^+_{Bones})}$$

$$= \frac{0.80}{0.80 + \left(\frac{0.85}{0.15}\right) * 0.20}$$

$$t^1_{Bones} = 0.4138$$

(5.3) $$f^1_{Bones} = 1 - P\left(Rel^-_{Bones} | Rep_{\neg Rel^-_{Bones}}\right) = 1 - \frac{P(Rel^-_{Bones})}{P(Rel^-_{Bones}) + \left(\frac{t_{study}}{f_{study}}\right) * P(\neg Rel^-_{Bones})}$$

$$= 1 - \frac{0.80}{0.80 + \left(\frac{0.85}{0.15}\right) * 0.20}$$

$$f^1_{Bones} = 0.5862$$

Using these new reliability values, if we now recalculate the confidence that Bones should have that his diagnosis is correct, we see that he should definitely ask some other doctor to recheck his diagnosis.

(5.5) $$\frac{P(P_{diagnosis\ X})}{P(P_{diagnosis\ X}) + \frac{q^1_{Bones}}{p^1_{Bones}} * P(\neg P_{diagnosis\ X})} = \frac{0.5}{0.5 + \frac{0.5862}{0.4138} * 0.5} = 0.4138$$

The important fact is not just that this analysis goes hand in hand with our intuition, but also that it is analogous to the analysis of the undercutting defeater.

5.3 Probabilistic Reports

Let us now take a closer look at another case from Christensen (2010), where the report seems to be somehow different.

C.5.2 (The Experiment): Mr. Spock is asked by Dr. Leonard "Bones" McCoy to be the subject in an experiment. While Mr. Spock is drinking some of the water that was offered him, he is listening to Bones' description of the experiment. Bones tells him that after a practice question, Mr. Spock will be administered a drug and will have to answer some further questions. The drug has been shown to degrade people's performance sharply in just this type of task. In fact, the results of some studies show that 80% of the people are unable to give the right answer after taking the drug, but would not recognize this malfunction by themselves and are still as confident as before. Mr. Spock is very interested and therefore immediately readies to answer the first question:

> Suppose that all bulls are fierce and Ferdinand is not a fierce bull. Which of the following must be true?
>
> (a) Ferdinand is fierce.
>
> (b) Ferdinand is not fierce.
>
> (c) Ferdinand is a bull.
>
> (d) Ferdinand is not a bull.

Mr. Spock becomes extremely confident that the answer is that only (b) must be true, and tells Bones so. Bones just smiles and tells Mr. Spock that the drug was already in the water that he just drank.

As in the later interpretation of C.4.2, we have a report on anti-reliability because people under the influence of the drug would say about what is the case that it is not, and about what is not the case that it is. Therefore, we could act like above. But there is an important difference. Bones is not merely saying that those who took the drug are anti-reliable; he is saying that 80% of those who took the drug show those effects. Therefore, we will have to think about how to manage such probabilistic reports.[54]

[54] For the following argumentation, see Jeffrey (1965), who develops the so-called *Jeffrey conditionalization*.

Let us say that one is giving the report that p is the case. Then, the influence on one's confidence in p changes as shown in 4.5, which means that

(5.6) $$\frac{P(p)}{P(p)+\frac{f}{t}*P(\neg p)}$$

where f and t are the reliability parameters of this source. If those sources will no longer say that p is the case, but that $\neg p$ is the case, our confidence that p is the case should change as we showed in 4.10. Therefore, our new level of confidence would be calculated by

(5.7) $$\frac{P(p)}{P(p)+\frac{t}{f}*P(\neg p)}$$

But now, what if the source S is saying that she herself is only confident by some level $\tilde{n}$ that p is the case, where $\tilde{n} = P_S(p)$? Then, what such a source is saying is that it is either that p is the case or that $\neg p$ is the case, where the first option has a probability of $\tilde{n}$ and the second of $1 - \tilde{n}$. Here, we have to count the effect of the positive report that p is true in relation to $\tilde{n}$, and the effect of the negative report that $\neg p$ is true in relation to $1 - \tilde{n}$. Let us call the effect of some positive report Δ^+, and the effect of some negative report Δ^-.

D.5.1 (Absolute Change): Δ^+, the *absolute positive change*, and Δ^- the *absolute negative change* indicate the positive or negative change between one's prior probability level to one post probability level after a positive or a negative report.

$$\Delta^+ := \frac{P^0(p)}{P^0(p) + \frac{f}{t} * P^0(\neg p)} - P^0(p)$$

$$\Delta^- := \frac{P^0(p)}{P^0(p) + \frac{t}{f} * P^0(\neg p)} - P^0(p)$$

By applying D.5.1 we can now say that one's post probability level after a positive report has to be the level of his prior confidence plus the corresponding absolute positive change. Analogous to this, one's post probability level after a negative report has to be the level of his prior confidence plus the corresponding absolute negative change. This means:

(5.8) $P\left(p\middle|Rep_p\right) = P^{1+}(p) = P^0(p) + \Delta^+$

(5.9) $P\left(p\middle|Rep_{\neg p}\right) = P^{1-}(p) = P^0(p) + \Delta^-$

As stated above, to get the report from some source S, that S has the confidence level $\tilde{n}$, has both, a positive and a negative effect on one's post probability level, the prior probability $\mathrm{P}^{1*}(\mathrm{p})$ after such a probabilistic report must also account for both of those effects relative to the confidence level. Therefore, we get:

(5.10) $P\left(p\middle|Rep_p^{\tilde{n}}\right) = \mathrm{P}^{1*}(\mathrm{p}) = \mathrm{P}^0(\mathrm{p}) + \Delta^+ * \tilde{n} + \Delta^- * (1 - \tilde{n})$[55]

We now see that the absolute positive change applies only if the source himself is absolutely confident—i.e. if $\tilde{n} = 1$—because under such circumstances $P^{1*}(p) = P^{1+}(p)$. If, on the other hand, the source is absolutely confident that p is not the case and, therefore, $\tilde{n} = 0$, we get $P^{1*}(p) = P^{1-}(p)$, which indicates the absolute negative change. This means that 5.8 and 5.9 are just special cases of 5.10.

These thoughts can now be applied on C.5.2. But we have to be careful because Bones is not just saying that Mr. Spock is anti-reliable with the confidence of $\tilde{n} = 0.8$. If he were, he would be expressing that he thinks Mr. Spock is reliable by a probability of $1 - \tilde{n} = 0.2$ in the strong sense of D.4. 5. Bones is saying rather that he thinks Mr. Spock is under the influence of the drug with the probability of $\tilde{n}$. But if he has not come under the influence, his reliability level is not Rel, but would stay unchanged This means

[55] See Jeffrey (1965: 169).

that the absolute positive effect of Bones' report is the absolute negative effect of the report of Mr. Spock's anti-reliability, where the absolute negative effect would be no change at all, and, therefore, $\Delta^- = 0$.

To analyze the influence on Mr. Spock, we must first consider some parameters. To keep it simple, we take for Bones the same values as above; thus, $t_{Bones} = 0.8$ and $f_{Bones} = 0.2$. Furthermore, we say that Mr. Spock is as good in answering such questions as Bones is in his field. Therefore, Mr. Spock has the same reliability parameters and we can say that $t_{Spock} = 0.8$ and $f_{Spcok} = 0.2$. We already know the absolute negative change. The question now is what is the absolute positive change—i.e. what if the drug works? Here, we can calculate the absolute positive change analogous to 4.26–4.29 and 5.8–5.10.

(5.11) $$\Delta_{\mathrm{t}}^{+} = t_{Spcok}^{1} - t_{spcok} = \frac{P\left(Rel_{Spcok}^{+}\right)}{P\left(Rel_{Spcok}^{+}\right) + \left(\frac{t_{Bones}}{t_{Bones}}\right) * P\left(\neg Rel_{Spcok}^{+}\right)} - t_{Spock}$$

$$= \frac{0.8}{0.8 + \left(\frac{0.8}{0.2}\right) * 0.2} - 0.8$$

$$\Delta_{\mathrm{t}}^{+} = -0.3$$

(5.12) $$\Delta_{\mathrm{f}}^{+} = f_{Spcok}^{1} - f_{spcok} = 1 - \frac{P\left(Rel_{Spcok}^{-}\right)}{P\left(Rel_{Spcok}^{-}\right) + \left(\frac{t_{Bones}}{f_{Bones}}\right) * P\left(\neg Rel_{Spcok}^{-}\right)} - f_{Spock}$$

$$= 1 - \frac{0.8}{0.8 + \left(\frac{0.8}{0.2}\right) * 0.2} - 0.2$$

$$\Delta_{\mathrm{f}}^{+} = 0.3$$

We can now insert those values in 5.10 and get the following:

(5.13) $$t_{Spcok}^{1*} = t_{Spock} + \Delta_{\mathrm{t}}^{+} * \tilde{\mathrm{n}} + \Delta_{\mathrm{t}}^{-} * (1 - \tilde{\mathrm{n}}) = 0.8 + (-0.3) * 0.8 + 0 * 0.2$$

$$t_{Spcok}^{1*} = 0.56$$

(5.14) $$f_{Spcok}^{1*} = f_{Spock} + \Delta_{\mathrm{f}}^{+} * \tilde{\mathrm{n}} + \Delta_{\mathrm{f}}^{-} * (1 - \tilde{\mathrm{n}}) = 0.2 + 0.3 * 0.8 + 0 * 0.2$$

$$f_{Spcok}^{1*} = 0.44$$

Let us now compare the confidence that Mr. Spock should have before receiving Bones' information about the drug, and then after Bones' confession. Since Spock has no previous knowledge about fierce bulls, we can set the prior probability $P(p_b) = 0.5$. On inserting these values in 4.5 we get

(5.15) $$P^1(p_b) = \frac{P(p_b)}{P(p_b)+\left(\frac{f_{Spock}}{t_{Spock}}\right)*P(\neg p_b)} = \frac{0.5}{0.5+\frac{0.2}{0.8}*0.5} = 0.8$$

But after being informed about the drug, Mr. Spock has to recalculate his confidence as follows.

(5.16) $$P^2(p_b) = \frac{P(p_b)}{P(p_b)+\left(\frac{f^{1*}_{Spcok}}{t^{1*}_{Spcok}}\right)*P(\neg p_b)} = \frac{0.5}{0.5+\frac{0.44}{0.56}*0.5} = 0.56$$

We see again, as in C.5.1, a change in our information about the source. Now, Mr. Spock himself causes a revaluation of the reliability of the source. Hence, he has to lower his confidence after receiving some negative HOE about his source, and this is the same reason why one has to lower one's confidence in cases involving the undercutting defeater.

6

Bayesian Epistemology and Disagreement

Christensen (2010) himself argues that peer disagreement is HOE. In the previous two chapters, I established a way to interpret cases of HOE by using the tools of Bayesian epistemology. Now, I will take this analysis and apply it on situations of peer disagreement. The result will not be an EWV, but can be interpreted as a version of the TEV.

First, we must consider what information one receives through peer disagreement and if it really is HOE. Here, I will work with the Jeffrey conditionalization since it is, as said above, the general cases besides the two special of absolute confidence of the source in some proposition or in its negation. I will use this information as I used the information from Christensen's HOE. The results will not be an EWV, where we have the situation of agreement after taking the HOE provided by the disagreement into account. It will be more like the TEV puts it—both parties take the HOE seriously, and lower their trust in one's evidence, but this does not necessarily lead to agreement.

The end of this chapter will visualize the results of such an analysis of disagreement. This visualization will be intuitively correct. But it also includes consequences that seem odd. I will try to explain the oddities to the best of my abilities, but at the end, everyone must decide on their own whether this explanation was successful. After all, this is always the risk of a formal analysis in the field of philosophy. One can only try to formalize some intuitive starting point as well as possible. But if the analysis leads one to an odd end, he only has two possibilities, given that the analysis was formally correct: change the starting point of the interpretation and look again, or deal with the oddity. I think the

latter is in some ways the surplus of a formal interpretation, since such an approach has the power to challenge our intuitions.

6.1 Peer Disagreement as Information

First, we must think about what the information of peer disagreement is, that must take into account to revise the degree of belief. Therefore, we use the definition D.2.3 of peer disagreement. By this definition, we know that in cases of peer disagreement, both agents are evidentially equal. This is an important point in peer disagreement because it rules out that the report of one's peer is just evidence for or against the proposition p about which they are disagreeing. If that were the case, peer disagreement would just be a situation of a rebutting defeater, where one gets additional evidence regarding whether p or $\neg p$ is true. That there is evidential equality, therefore, indicates that the opponent's opinion is not information about whether p is true but rather about the body of evidence itself. In other words, I agree with Christensen on this point. Therefore, an interpretation of peer disagreement as HOE seems to me to fit very well. But the difference is that I also see peer disagreement as a UD. As stated above, there is no fundamental difference between the two. Therefore, in the following I will refer to peer disagreement as HOE. The question now is what HOE is peer disagreement exactly. Earlier, we saw two fundamentally different cases with C.5.1 and C.5.2. The first, C.5.1, was parallel to C.4.2, the case where we acquired some negative information about our source. The case C.5.2 did nearly the same thing, but here we got a probabilistic report. Since C.5.2 is the more general case, as said above, I will use it as the reference to interpret cases of peer disagreement. Therefore, let us take a look at a case of peer disagreement like C.2.1, but this time, in terms of degrees of belief.

C.6.1 (Warp Drive III): After an attack, the warp drive of the USS Enterprise is not working. Since the rest of the engineer corps died during the attack, Mr. Spock is now in the engine room to help Montgomery "Scotty" Scott. Although Scotty is the chief engineer on board the Enterprise, they are both aware that Mr. Spock knows about the

warp drive as much as he does. After evaluating all the relevant facts about the damage, they are talking about the possibility to fix the warp drive. But although they have equal knowledge about the damage, and one is as good as the other in evaluating the possibility to fix the warp drive, Mr. Spock thinks that they will be able to fix the warp drive with a degree of belief of 0.6, while Scotty's degree of belief is only 0.3.

For the proposition that the warp drive can be fixed, we again take the term p_{WD}. The degrees of belief of Mr. Spock and Scotty are $P_{Spock}(p_{WD})$ and $P_{Scotty}(p_{WD})$. From C.6.1, we only know that Mr. Spock and Scotty are also cognitive equals, but we now have to be more concrete. Therefore, we have to define their true positive rate and their false negative rate.

(6.1) $t_{Spock} = t_{Scotty} \coloneqq 0.90$

(6.2) $f_{Spock} = f_{Scotty} \coloneqq 0.20$[56]

These values indicate how credible both, Mr. Spock and Scotty, are in evaluating some body of evidence about the warp drive.

Since we also assume that Mr. Spock and Scotty are rational agents, we know that they formed their degrees of belief through Bayesian rules. If we consider the Jeffrey conditioning we know that, for example, Mr. Spock's degree of belief was calculated by

(6.3) $P_{Spock}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) = P^0_{Spock}(p_{WD}) + \Delta^{+}_{\text{Spock}} * \tilde{n} + \Delta^{-}_{\text{Spock}} * (1 - \tilde{n}),$

Here, Δ^{+} and Δ^{-} as defined in D.5.1 can be calculated for Mr. Spock as follows.

(6.4) $$\Delta^{p_{WD}+}_{\text{Spock}} = \frac{P^0_{Spock}(p_{WD})}{P^0_{Spock}(p_{WD}) + \frac{f_{Spock}}{t_{Spock}} * P^0_{Spock}(\neg p_{WD})} - P^0_{Spock}(p_{WD})$$

(6.5) $$\Delta^{p_{WD}-}_{\text{Spock}} = \frac{P^0_{Spock}(p_{WD})}{P^0_{Spock}(p_{WD}) + \frac{t_{Spock}}{f_{Spock}} * P^0_{Spock}(\neg p_{WD})} - P^0_{Spock}(p_{WD})$$

[56] We worked with other values for these two agents, but such values are always topic related, and since we changed the topic, a change in t and f occur as a consequence.

The question now is, of course, what is the rational prior $P^0_{Spock}(p_{WD})$? As argued above, we are here talking about a situation where Mr. Spock does not have evidence for or against p_{WD}. Therefore, it seems irrational to prefer one possibility. Since this holds for Mr. Spock and Scotty, let us define the priors with

(6.6) $$P^0_{Spock}(p_{WD}) = P^0_{Scotty}(p_{WD}) \coloneqq 0.5$$

Using this value, we can calculate $\Delta^{+}_{\text{Spock}}$ and $\Delta^{-}_{\text{Spock}}$ as follows.

(6.7) $$\Delta^{p_{WD}+}_{\text{Spock}} = \frac{0.5}{0.5+\frac{0.3}{0.8}*(1-0.5)} - 0.5 = 0.318$$

(6.8) $$\Delta^{p_{WD}-}_{\text{Spock}} = \frac{0.5}{0.5+\frac{0.8}{0.05}*(1-0.5)} - 0.5 = -0.318$$

But now, if we consider (6.1) and (6.2), the cognitive equality, as well as (6.6), we can see that

(6.9) $$\Delta^{p_{WD}+}_{\text{Spock}} = \Delta^{p_{WD}+}_{\text{Spock}},$$

and

(6.10) $$\Delta^{p_{WD}-}_{\text{Spock}} = \Delta^{p_{WD}-}_{\text{Spock}},$$

since all relevant values in (6.4) and (6.5) are the same for Mr. Spock and Scotty.

With this in mind, we can look again to (6.3), where we can see that the only variable that is still unknown is ñ. This is also the only variable that can explain the different levels of degree. Disagreements between peers are, therefore, HOE concerning a concrete value of my body of evidence, ñ. Still, taking Mr. Spock as an example, we can calculate his personal value for this variable by

(6.11) $$P_{Spock}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) = P^0_{Spock}(p_{WD}) + \Delta^{p_{WD}+}_{\text{Spock}} * \tilde{n} + \Delta^{p_{WD}-}_{\text{Spock}} * (1-\tilde{n})$$

$$P_{Spock}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) - P^0_{Spock}(p_{WD}) = \Delta^{p_{WD}+}_{\text{Spock}} * \tilde{n} + \Delta^{p_{WD}-}_{\text{Spock}} - \Delta^{p_{WD}-}_{\text{Spock}} * \tilde{n}$$

$$P_{Spock}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) - P^0_{Spock}(p_{WD}) - \Delta^{p_{WD}-}_{\text{Spock}} = \left(\Delta^{p_{WD}+}_{\text{Spock}} - \Delta^{p_{WD}-}_{\text{Spock}}\right) * \tilde{n}$$

$$\tilde{n} = \frac{P_{Spock}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) - P^{0}_{Spock}(p_{WD}) - \Delta^{p_{WD}-}_{\text{Spock}}}{\Delta^{p_{WD}+}_{\text{Spock}} - \Delta^{p_{WD}-}_{\text{Spock}}}$$

Since we already know from C.6.1 that $P_{Spock}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) = 0.6$, we now only have to insert the values for all the parameters and get

(6.12) $\tilde{n}_{Spock} = \frac{0.6-0.5-(-0.318)}{0.318-(-0.318)} = 0.657$

Everything we said also holds for Scotty. Therefore, we can calculate his personal value for $\tilde{n}$ in a similar way to Mr. Spock's. The only difference is that since Mr. Spock and Scotty are in the situation of disagreement about p_{WD}, Scotty's value for $P_{Scotty}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right)$ differs from that of Mr. Spock. As we can see in C.6.1, Scotty's confidence in the possibility to fix the warp drive, after the evaluation of the information is 0.3. Therefore, we have $P_{Scotty}\left(p_{WD}\middle|E^{\tilde{n}}_{p_{WD}}\right) = 0.3$. Taking this into account, we can insert the values for Scotty into the formula 6.11 and get

(6.13) $\tilde{n}_{Scotty} = \frac{0.3-0.5-(-0.318)}{0.318-(-0.318)} = 0.186.$

The reason for the disagreement between Mr. Spock and Scotty is, therefore, their different evaluations of the evidence. Thus, we can evaluate the information that one receives by peer disagreement as information about exactly this parameter regarding one's evidence. If Mr. Spock and Scotty now want to take this information from their peer disagreement into account, they have to consider the implied assertions of $\tilde{n}_{Spock}$ and $\tilde{n}_{Scotty}$ as a report from their peers and to evaluate this report as we did in cases of HOE.

6.2 Evaluating the Information of Peer Disagreement

The rational influence of such an information of peer disagreement can be calculated by the means of BE. The prior values of one's degree of belief that one's body of evidence indicates that p is $\tilde{n}$ can be updated by the Jeffrey rule of conditionalization, as we did in

chapter 5 for case C.5.2. Here, we have a propositional report about ñ. Let us call the proposition that ñ has the value 1 $p_{ñ}$. Then, we get the following two equations for Mr. Spock and Scotty.

(6.14) $P_{Spock}\left(p_{ñ}\middle|Rep_{p_{ñ}}^{ñ_{Scotty}}\right) = \mathrm{P}^{0}_{\mathrm{Spock}}(p_{ñ}) + \Delta^{p_{ñ}+}_{\mathrm{Spock}} * ñ_{Scotty} + \Delta^{p_{ñ}-}_{\mathrm{Spock}} * (1 - ñ_{Scotty})$

and

(6.15) $P_{Scotty}\left(p_{ñ}\middle|Rep_{p_{ñ}}^{ñ_{Spock}}\right) = \mathrm{P}^{0}_{\mathrm{Scotty}}(p_{ñ}) + \Delta^{p_{ñ}+}_{\mathrm{Scotty}} * ñ_{Spock} + \Delta^{p_{ñ}-}_{\mathrm{Scotty}} * (1 - ñ_{Spock})$,

where $\mathrm{P}^{0}_{\mathrm{Spock}}(p_{ñ})$ stands for $ñ_{Spock}$ and $\mathrm{P}^{0}_{\mathrm{Scotty}}(p_{ñ})$ for $ñ_{Scotty}$. Now, we have to think about the calculation of the absolute positive and absolute negative change for Mr. Spock and Scotty.

The general formulas for those two values are given in D.5.1 as

$$\Delta^{+} := \frac{P^{0}(p)}{P^{0}(p) + \frac{f}{t} * P^{0}(\neg p)} - P^{0}(p)$$

$$\Delta^{-} := \frac{P^{0}(p)}{P^{0}(p) + \frac{t}{f} * P^{0}(\neg p)} - P^{0}(p)$$

This means that we have to think about the priors and the values t and f. The priors are, as stated above, the values for ñ from Mr. Spock and Scotty before the disagreement. But what are the values for t and f? Since the report is a report about the evaluation of evidence, and we defined in 6.1 and 6.2 the credibility of Mr. Spock and Scotty in evaluating a body of evidence, those are the values we need. Therefore, we can rewrite the formula for the absolute positive and negative rate as an example for Mr. Spock and get

(6.16) $\Delta^{p_{ñ}+}_{\mathrm{Spock}} = \frac{ñ_{Spock}}{ñ_{Spock} + \frac{f_{Scotty}}{t_{Scotty}} * (1 - ñ_{Spock})} - ñ_{Spock}$

(6.17) $\Delta^{p_{ñ}-}_{\mathrm{Spock}} = \frac{ñ_{Spock}}{ñ_{Spock} + \frac{t_{Scotty}}{f_{Scotty}} * (1 - ñ_{Spock})} - ñ_{Spock}$

If we now insert Mr. Spock's values into 6.16 and 6.17 and to the analogue for Scotty, we get the following four parameters

(6.18) $\Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}+} = \frac{0.657}{0.657+\frac{0.30}{0.90}*(1-0.657)} - 0.657 = 0.239$

(6.19) $\Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}-} = \frac{0.657}{0.657+\frac{0.90}{0.30}*(1-0.657)} - 0.657 = -0.358$

(6.20) $\Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}+} = \frac{0.186}{0.186+\frac{0.30}{0.90}*(1-0.186)} - 0.186 = 0.320$

(6.21) $\Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}-} = \frac{0.186}{0.186+\frac{0.90}{0.30}*(1-0186)} - 0.186 = -0.137$

On reinserting this into 6.14 and 6.15, we get the post values for $\tilde{\text{n}}$.

(6.22) $\tilde{\text{n}}_{Spock}^{1} = P_{Spock}\left(p_{\tilde{\text{n}}}\middle|Rep_{p_{\tilde{\text{n}}}}^{\tilde{\text{n}}_{Scotty}}\right) = \text{P}_{\text{Spock}}^{0}\left(p_{\tilde{\text{n}}}\right) + \Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}+} * \tilde{\text{n}}_{Scotty} + \Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}-} * (1 - \tilde{\text{n}}_{Scotty})$

$\tilde{\text{n}}_{Spock}^{1} = 0.657 + 0.239 * 0.186 + (-0.358) * (1 - 0.186)$

$\tilde{\text{n}}_{Spock}^{1} = 0.409$

If we do the same for Scotty, we get

(6.23) $\tilde{\text{n}}_{Scotty}^{1} = 0.349$

The results seem intuitively true. After the report from Scotty that the body of evidence does not favor p_{WD} as much as Mr. Spock thinks it does, he should lower his trust in the body of evidence. The inverse should hold for Scotty. The results from 6.22 and 6.23 show exactly this. Mr. Spock lowers his confidence while Scotty raises his confidence. We get their new level of confidence in p_{WD} on reinserting these new levels of $\tilde{\text{n}}_{Spock}^{1}$ and $\tilde{\text{n}}_{Scotty}^{1}$ into 6.3. Then, we get the two formulas

(6.24) $P_{Spock}^{1}\left(p_{WD}\middle|E_{p_{WD}}^{\tilde{\text{n}}_{Spock}^{1}}\right) = P_{Spock}^{0}(p_{WD}) + \Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}+} * \tilde{\text{n}}_{Spock}^{1} + \Delta_{\text{Spock}}^{p_{\tilde{\text{n}}}-} * (1 - \tilde{\text{n}}_{Spock}^{1})$,

and

(6.25) $P_{Scotty}^{1}\left(p_{WD}\middle|E_{p_{WD}}^{\tilde{\text{n}}_{Scotty}^{1}}\right) = P_{Scotty}^{0}(p_{WD}) + \Delta_{\text{Scotty}}^{p_{\tilde{\text{n}}}+} * \tilde{\text{n}}_{Spock}^{1} + \Delta_{\text{Scotty}}^{p_{\tilde{\text{n}}}-} * (1 - \tilde{\text{n}}_{Spock}^{1})$.

The value for $P^0_{Spock}(p_{WD})$ and $P^0_{Scotty}(p_{WD})$ is again 0.5. The values for the absolute positive and negative change also remain the same, since 6.4 and 6.5 make no reference to the re-evaluated term ñ. Thus, we have all the parameters from 6.24 and 6.25, and can calculate Mr. Spock's and Scotty's post level of confidence in p_{WD} as follows.

(6.26) $$P^1_{Spock}\left(p_{WD}\middle|E^{\tilde{n}^1_{Spock}}_{p_{WD}}\right) = 0.5 + 0.318 * 0.409 + (-0.318) * (1 - 0.409) = 0.442$$

(6.27) $$P^1_{Scotty}\left(p_{WD}\middle|E^{\tilde{n}^1_{Scotty}}_{p_{WD}}\right) = 0.5 + 0.318 * 0.349 + 0.318 * (1 - 0.349) = 0.404$$

These results are interesting. We see that both parties have to lower their confidence in the light of peer disagreement. But taking the claim that peer disagreement is HOE seriously does not lead to the EWV since the levels of confidence are not the same even after evaluating the information. Bringing together the FOE, the starting point of the disagreement, and the HOE, the disagreement itself, as the TEV suggests, leads to the results in 6.26 and 6.27.

A further interesting finding of this analysis is that the fact that the opponents are cognitive equals was never really important. If Mr. Spock and Scotty were not evidentially equal, the analysis would not work, since the information in this situation would be something else. Mr. Spock's and Scotty's being cognitive equals, however, made it easier to calculate the results, but was not really necessary. We could have done the same if Scotty were a cognitive superior to Mr. Spock or vice versa. In the next section, I will present some visualizations of peer disagreement, which will show the influence of different values of t and f for the opponents.

6.3 Visualizing the Influence of Peer Disagreement

As mentioned above, cognitive equality was not really necessary for the calculation of the post level of confidence. Therefore, let us see how a change in Scotty's reliability values, t_{Scotty} and f_{Scotty}, would change the results. The values for Mr. Spock are the same as above; hence, we already know that $t_{Spock} = 0.90$ and $f_{Spock} = 0.20$. The prior value of

confidence is also the same, thus $P^0_{Spock}(p_{WD}) = P^0_{Scotty}(p_{WD}) \coloneqq 0.5$, and from case C.6.1, we still get $P_{Spock}\left(p_{\mathrm{WD}}\middle|Rep^{\tilde{\mathrm{n}}}_{p_{\mathrm{WD}}}\right) = 0.6$ and $P_{Scotty}\left(p_{\tilde{\mathrm{n}}}\middle|Rep^{\tilde{\mathrm{n}}}_{p_{\mathrm{WD}}}\right) = 0.3$, where the terms $P_{Spock}\left(p_{\tilde{\mathrm{n}}}\middle|Rep^{\tilde{\mathrm{n}}}_{p_{\mathrm{WD}}}\right)$ and $P_{Scotty}\left(p_{\mathrm{WD}}\middle|Rep^{\tilde{\mathrm{n}}}_{p_{\mathrm{WD}}}\right)$ stand for the level of confidence towards p_{WD} of Mr. Spock and Scotty after evaluating the body of evidence that both possess. If we now want to construct a two-dimensional graph, we need a variable that includes both of the reliability values of Scotty. Such a variable can be found in D.4.4. with $r \coloneqq 1 - \frac{f}{t}$. We additionally define $\overline{r} \coloneqq 1 - r$. Using this variable, we can define two formulas with the variable $\overline{r}$ as the argument and the post level of confidence of Mr. Spock and Scotty as the result. We can start with Scotty's absolute positive and negative change given some body of evidence concerning p_{WD}.

(6.28) $$\Delta^{p_{WD}+}_{\mathrm{Scotty}}(\overline{r}) = \frac{P^0_{Scotty}(p_{WD})}{P^0_{Scotty}(p_{WD}) + \overline{r} * P^0_{Scotty}(\neg p_{WD})} - P^0_{Scotty}(p_{WD})$$

(6.29) $$\Delta^{p_{WD}-}_{\mathrm{Scotty}}(\overline{r}) = \frac{P^0_{Scotty}(p_{WD})}{P^0_{Scotty}(p_{WD}) + \frac{1}{\overline{r}} * P^0_{Scotty}(\neg p_{WD})} - P^0_{Scotty}(p_{WD})$$

By following 6.11, we also get a formula for $\tilde{\mathrm{n}}_{Scotty}$ where $\overline{r}$ is the argument.

(6.30) $$\tilde{\mathrm{n}}_{Scotty}(\overline{r}) = \frac{P_{Scotty}\left(p_{WD}\middle|E^{\tilde{\mathrm{n}}}_{p_{WD}}\right) - P^0_{Scotty}(p_{WD}) - \Delta^{p_{WD}-}_{\mathrm{Scotty}}(\overline{r})}{\Delta^{p_{WD}+}_{\mathrm{Scotty}}(\overline{r}) - \Delta^{p_{WD}-}_{\mathrm{Scotty}}(\overline{r})}$$

We additionally get the values for the absolute positive and negative change concerning some information about the body of evidence from Mr. Spock and Scotty, following 6.16 and 6.17. For Mr. Spock, we only have to insert $\overline{r}$, since all the other values refer only to unchanged variables.

(6.31) $$\Delta^{p_{\tilde{\mathrm{n}}}+}_{\mathrm{Spock}}(\overline{r}) = \frac{\tilde{\mathrm{n}}_{Spock}}{\tilde{\mathrm{n}}_{Spock} + \overline{r} * (1 - \tilde{\mathrm{n}}_{Spock})} - \tilde{\mathrm{n}}_{Spock}$$

(6.32) $$\Delta^{p_{\tilde{\mathrm{n}}}-}_{\mathrm{Spock}}(\overline{r}) = \frac{\tilde{\mathrm{n}}_{Spock}}{\tilde{\mathrm{n}}_{Spock} + \frac{1}{\overline{r}} * (1 - \tilde{\mathrm{n}}_{Spock})} - \tilde{\mathrm{n}}_{Spock}$$

For Scotty, we get

(6.32) $$\Delta^{p_{\tilde{n}}+}_{\text{Scotty}}(\overline{r}) = \frac{\tilde{n}_{Scotty}(\overline{r})}{\tilde{n}_{Scotty(\overline{r})}+\frac{f_{Spock}}{t_{Spock}}*(1-\tilde{n}_{Scotty}(\overline{r}))} - \tilde{n}_{Scotty}(\overline{r})$$

(6.34) $$\Delta^{p_{\tilde{n}}-}_{\text{Scotyy}}(\overline{r}) = \frac{\tilde{n}_{Scotty}(\overline{r})}{\tilde{n}_{Scotty(\overline{r})}+\frac{t_{Spock}}{f_{Spock}}*(1-\tilde{n}_{Scotty}(\overline{r}))} - \tilde{n}_{Scotty}(\overline{r})$$

From 6.22, we saw how these values influence the posterior evaluation of the information. If we insert the parameters from 6.31–6.34, we again get formulas with $\overline{r}$ as an argument.

(6.35) $$\tilde{n}^1_{Spock}(\overline{r}) = \tilde{n}_{Spock} + \Delta^{p_{\tilde{n}}+}_{\text{Spock}}(\overline{r}) * \tilde{n}_{Scotty}(\overline{r}) + \Delta^{p_{\tilde{n}}-}_{\text{Spock}}(\overline{r}) * (1 - \tilde{n}_{Scotty}(\overline{r}))$$

(6.36) $$\tilde{n}^1_{Scotty}(\overline{r}) = \tilde{n}_{Scotty}(\overline{r}) + \Delta^{p_{\tilde{n}}+}_{\text{Scotty}}(\overline{r}) * \tilde{n}_{Spock} + \Delta^{p_{\tilde{n}}-}_{\text{Scotty}}(\overline{r}) * (1 - \tilde{n}_{Spock})$$

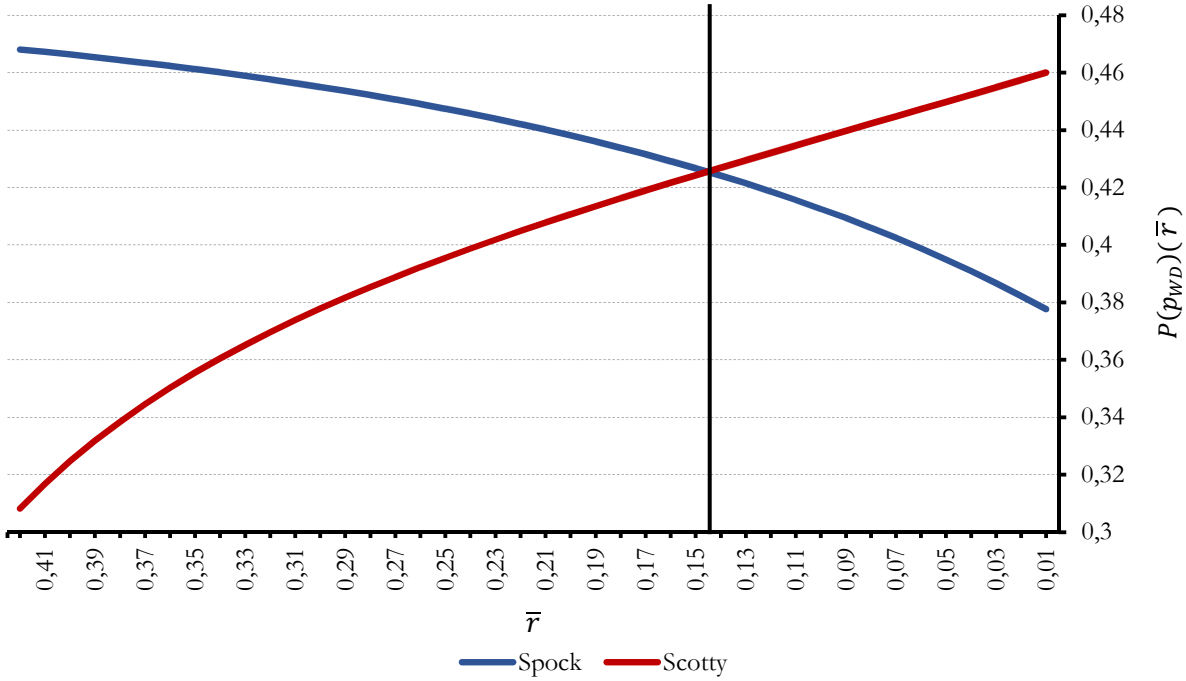
These values can be reinserted in the calculation of the rational level of confidence as we did in 6.25 and 6.26, and we get the formulas for calculating the post level of confidence with the parameter $\overline{r}$ for Scotty's reliability as an argument.

(6.37) $$P^1_{Spock}\left(p_{WD}\middle|E^{\tilde{n}^1_{Spock}}_{p_{WD}}\right)(\overline{r}) = P^0_{Spock}(p_{WD}) + \Delta^{p_{WD}+}_{\text{Spock}} * \tilde{n}^1_{Spock}(\overline{r}) + \Delta^{p_{WD}-}_{\text{Spock}} * (1 - \tilde{n}^1_{Spock}(\overline{r}))$$

(6.38) $$P^1_{Scotty}\left(p_{WD}\middle|E^{\tilde{n}^1_{Scotty}}_{p_{WD}}\right)(\overline{r}) = P^0_{Scotty}(p_{WD}) + \Delta^{p_{WD}}_{\text{Scotty}}(\overline{r}) * \tilde{n}^1_{Spock}(\overline{r}) + \Delta^{p_{WD}}_{\text{Scotty}}(\overline{r}) * (1 - \tilde{n}^1_{Spock}(\overline{r}))$$

Using these two formulas, we can generate the following graph.

Fig. 1



This gives rise to the question of why the value for $\overline{r}$ starts at around 0.42 and not at 1. The value 1 would mean that Scotty is totally unreliable, but since we fixed his prior level of confidence towards p_{WD} at 0.3, he has to be at least as reliable that the value $\overline{r} \leq 0.42$. This is because the value for $\tilde{n}$ cannot be higher than 1 and lower than 0, since this would already mean that the body of evidence speaks absolutely in favor of p_{WD} and $\neg p_{WD}$ respectively. Therefore, if we want Scotty's level of confidence towards p_{WD} fixed at 0.3, his value $\overline{r}$ has to be such, that $1 \leq \tilde{n} \leq 0$.[57] But if $\overline{r} > 0.42$, the value for $\tilde{n}$ becomes less then 0 as we can easily see from 6.30, since 6.30 would lead to the following development of $\tilde{n}_{Scotty}$ depending on $\overline{r}$.

$\overline{r}$	0.35	0.36	0.37	0.38	0.39	0.40	0.41	0.42	0.43	0.44	0.45
$\tilde{n}_{Scotty}(\overline{r})$	0.085	0.075	0.065	0.055	0.044	$0.0\dot{3}$	0.022	0.010	-0.001	-0.014	-0.027

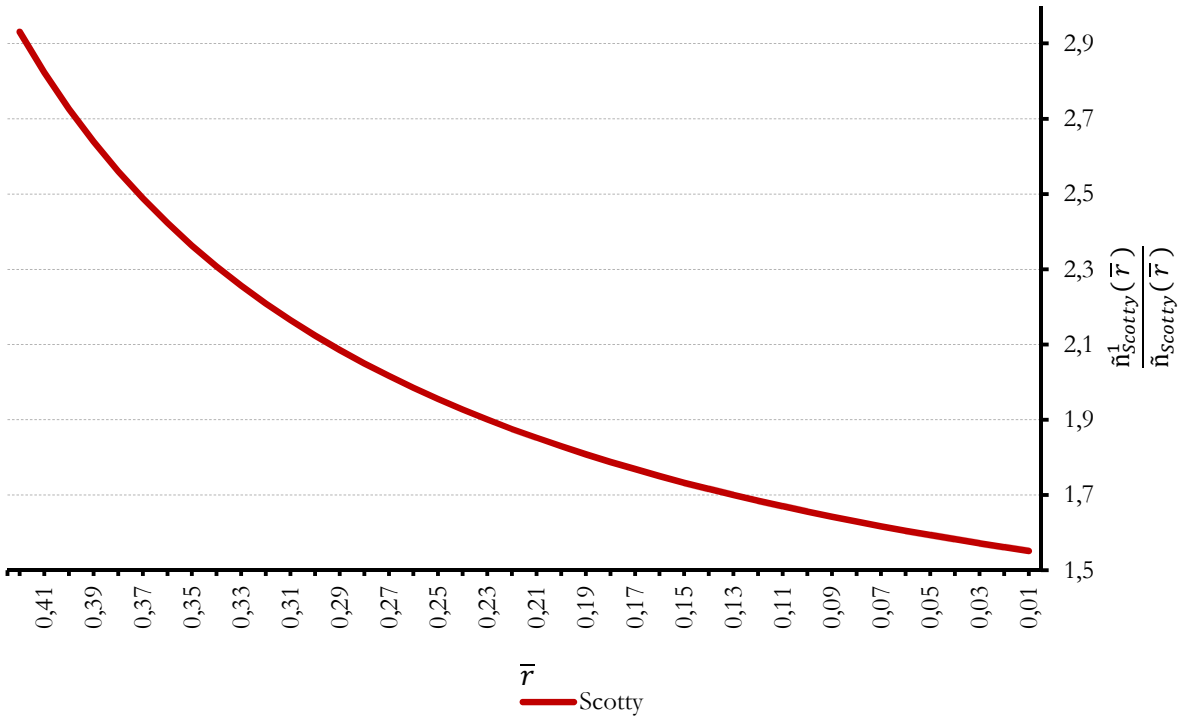
We can see that if Scotty becomes less reliable as the value $\overline{r} = 0.42$ represents, he cannot rationally hold on to a belief level of 0.3 without violating the constraints of $\tilde{n}$. Thus, given that 0.3 is Scotty's level of confidence, it includes that $\overline{r} \leq 0.42$.

[57] This can be seen by the Kolmogorov axiomatization in D.3.3.

Let us now interpret Fig. 1. The blue line describes Mr. Spock's confidence level regarding the possibility that the warp drive can be fixed, following the information of the disagreement with Scotty. Since Scotty becomes more and more reliable, Mr. Spock has to take Scotty's opinion more and more reliable. This effect is seen in the blue line. If Scotty is almost absolutely reliable, Mr. Spock nearly adopts his opinion. This is shown at the right end of the graph. We also see that there is a point where the information from the disagreement leads to agreement. This is not the case when Mr. Spock and Scotty are cognitive equals, but when Scotty's reliability parameter $\overline{r} \approx 0{,}145$. When Mr. Spock and Scotty are cognitive equals, it means that $t_{Spock} = t_{Scotty} = 0.90$ and $f_{Spock} = f_{Scotty} = 0.20$, and this parameter has the value $\overline{r}_{equal} = 0,\dot{2}$. Since the lower this value is, the more reliable Scotty is, it indicates that only if Scotty is more reliable than Mr. Spock the information of the disagreement does lead to agreement.

But there are two further effects of this analysis that are not as intuitive as the falling blue line. First, why is Scotty's red line rising—in other words, why does he give up more and more of his opinion while becoming more reliable? This effect emerges since we hold his prior level of confidence as fixed. Because of this, we change his evaluation of how much the body of evidence favors p_{WD}. This is interesting because if we look at the extent to which he changes his opinion regarding the body of evidence in the light of the disagreement, we see an effect that is represented in the following graph.

Fig. 2



We see here that in the case that Scotty is as unreliable as possible given his level of confidence, and therefore the value of $\overline{r}$ is as high as possiple, he has to raise the confidence that the body of evidence speaks in favour for p_{WD}about 3 times. Meanwhile, in the case that Scotty is very reliable, the value $\tilde{n}$ rises only around 1.5 times. This is exactly the effect that we would intuitively expect.

But there is still one effect left to discuss and that is the behavior of the lines in the situations where $\overline{r} < 0.145$. This result seems rather odd. Before the disagreement, Mr. Spock is more confident that p_{WD} is true than Scotty is, but if Scotty's reliability parameter fulfills $\overline{r} < 0.145$, then Mr. Spock is less confident after the disagreement that p_{WD} is true than is Scotty. The results before this are in consensus with our intuitions. Here, we see the effect that Mr. Spock and Scotty have lower and higher confidence respectively, which consequently brings their level of confidence closer. But after the point of the agreement, they somehow overshoot the mark. On the other side, it seems clear that in the case that one has a disagreement with a very reliable source, one should go very near his position. But if the one is a very reliable source too, then the opponent

should do the same, which would lead to exactly the result that the lines show in the situation of $\overline{r} < 0.145$.

7

Summary

The aim of this work was to develop a new view on the problems of disagreement emerging from Bayesian epistemology. Since disagreement is such a common event, we have strong intuitions concerning this topic. When I am disagreeing with my friends, I have to agree with Van Inwagen (2010) in that I do not want to take away their status of a cognitive equality. I am also unwilling to give up my doxastic attitude and would be more than happy if my friends would do so. In this work, I tried to take disagreement seriously—i.e. to think of disagreement as a sort of information. The EWV claims to do so, but here I miss the evaluation of the FOE. Just to take the symmetry of the disagreement into account seems insufficient to me. The level of reliability, even if equal, has to play a role. Moreover, the level of confidence should divide different cases of disagreement. Therefore, I thought that an account of the TEV would be more accurate. My interpretation of the TEV is that one has to take the HOE that disagreement is, use it to re-evaluate one's FOE, and then recalculate the rational level of confidence through this new FOE. To do so, I first argued that the FOE still matters in the situation of disagreement. I did this by showing that the difference between HOE and UD, as argued by Christensen, does not hold. Since he used this difference to argue that in the case where one has HOE, he has to *put aside* the FOE; hence, he loses this case by showing that there is no principal difference between HOE and UD. Here, I mentioned the difference between environmental UD and source UD. I think it would be interesting to work out this difference in detail and I hope to do so soon. But the main idea here was that UD and HOE are of the same type, just that HOE is a concept that can be phrased in terms of degrees of belief, while a defeater like UD is a concept that is phrased in

terms of a three-fold scheme of beliefs. My idea here is that UD is a HOE that is strong enough to change one's doxastic attitude away from belief. But to provide a detailed definition of this relationship, I would need an accurate translation from one concept of belief to the other. I think that there is a way to do this through an approach similar to the Dutch book argument, but this is still just a pie in the sky. Again, I hope that I will have more to present on this topic in the future. Such a translation could enable one to define UD as a special kind of HOE and the rebutting defeater as a special kind of FOE. At least, this is the idea that led to the notes of chapter 5.

In this work, I used the analysis from chapters 4 and 5 as a tool to analyze a case of disagreement. I interpreted disagreement as information about the parameter ñ of my body of evidence. This parameter is an indicator of how much my body of evidence speaks in favor of some propositions. To take disagreement as such information, the most important aspect is evidential equality. Only then does disagreement become pure HOE; otherwise, it would just be additional FOE, which I would get through some person. In such cases, disagreement is not HOE and therefore cannot be a UD, but it is FOE and should be interpreted like the rebutting defeater in C.4.2. The cognitive equality, on the other hand, is not a necessary condition to calculate the influence.

The results of chapter 6, where I also give a visualization of peer disagreement depending on the reliability of one of the opponents, are very interesting. On the one hand, they seem intuitively correct. They show how the influence should become stronger if the opponent becomes more reliable. The downside is that this result implies that the disagreement between very reliable peers is worse for the parties than is the disagreement of very unreliable agents. But I guess this result is not a new one, since Ginet (1980) had already discussed a similar problem.

On the other hand, the results also imply some oddity—namely, a change in the positions of the opponents. This means that in some situations, given the right level of confidence and reliability, the information of the disagreement can lead to a situation where the party that was more confident that some proposition p is true as was the opponent becomes, after the disagreement, the party that has less confidence in p than does the opponent. This can be seen at the right end of Fig.1. This is a result one might

not expect. But I was surprised by how many people thought such a result to be correct during discussions. Hence, I can only say that to me it seems odd, but not necessarily false.

This leads me to my final remark. I used objective Bayesian epistemology and worked with a prior of 0.5 in cases where agents know nothing about the truth or falsity of the proposition. It would be interesting to do an analogous analysis involving a different approach—namely, the approach presented in Kaplan (1996).[58] Kaplan tries to develop a position, where one lacks a concrete value for one's degree of confidence, but has a set of values. In this sense, the prior in cases where one knows nothing about the proposition is not a concrete value, but rather, the set $(0,1)$ of all possible values. But here I am in one of the minefields of Bayesian epistemology,[59] and since the aim of this work was not to solve these problems, I only used the basic tools of Bayesian epistemology for my analysis.

I think that the results of this analysis are at least worth discussing. I tried to work out a concrete example of a TEV that does not just say that in some cases one should lower one's confidence to a greater extent and in some other cases to a lesser extent. I tried to identify the important parameters, and work out why different changes are reasonable. I realize that a vague formulation of the TEV has the advantage that it is not as vulnerable since it is elusive. I definitively lost this advantage by applying Bayesian epistemology and working with concrete values of confidence, and especially with a formal approach to their dynamics. But I believe that the slightest chance of a new view at the debate around epistemology of disagreement is worth this risk.

[58] A discussion of the problem with sharp numeral values can be found in Joyce (2005: 156–158).
[59] For an overview of the problems of Bayesian epistemology, see Easwaran (2011b).

References

Bergmann, Michael (2009). 'Rational Disagreement after Full Disclosure', *Episteme* 6(3): 336–353 [doi:10.3366/E1742360009000756]

Bogardus, Tomas (2009). 'A Vindication of the Equal Weight View', *Episteme* 6(3): 324–335 [doi:10.3366/E1742360009000744]

Bovens, Luc and Hartman, Stephan (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Christensen, David (2007). 'Epistemology of Disagreement: The Good News', *Philosophical Review* 116(2): 187–217.

Christensen, David (2009a). 'Introduction: The Epistemology of Disagreement', *Episteme* 6(3): 231–232 [doi:10.3366/E1742360009000677]

Christensen, David (2009b). 'Disagreement as Evidence: The Epistemology of Controversy', *Philosophy Compass* 4(5): 756–767.

Christensen, David (2010). 'Higher-Order Evidence', *Philosophy and Phenomenological Research* 81(1): 185–215.

De Finetti, Bruno (1937). 'La prevision: ses lois logiques, ses sources subjectives.' *Annales de l'Institut Henri Poincare* 7: 1–68 (English translation 'Foresight: Its Logical Laws, Its Subjective Sources', in Henry E. Kyburg and Howard E.K Smokler (eds.), *Studies in Subjective Probability, Huntington* (1964), New York: Robert E. Krieger Publishing Co: 93–158).

Dennett, Daniel C. (1995). 'Intuition Pumps', in Brockman, J. (ed.), *The Third Culture: Beyond the Scientific Revolution*. New York: Simon & Schuster, 180–197.

Easwaran, Kenny (2011b). 'Bayesianism II', *Philosophy Compass* 6(5): 321–332.

Easwaran, Kenny (2013). 'Dutch Book Arguments', in Hal Pashler *et al.* (eds.), *The Encyclopedia of the Mind*. Ort: Verlag. 264–266.

Egan, Andy and Elga, Adam (2005). 'I Can't Believe I'm Stupid', *Philosophical Perspectives* 19(1): S. 77–93.

Elga, Adam (2007). 'Reflection and Disagreement', *Nous* 41(3): 487–502.

Feldman, Richard (2006). 'Epistemological Puzzles about Disagreement' in Stephen Hetherington (ed.), *Epistemology Futures*. Oxford: Oxford University Press (2006), 216–236.

Feldman, Richard (2007). 'Reasonable Religious Disagreement' in Louse Antony (ed.), *Philosophers without Gods: Meditations on Atheism and the Secular Life*. Oxford: Oxford University Press, 194–214.

Frances, Bryan (2010b). 'Disagreement.' in Sven Bernecker, Duncan Pritchard (eds.), *The Routledge Companion to Epistemology*. New York: Routledge, 63–74.

Ginet, Carl (1980). 'Knowing Less by Knowing More', *Midwest Studies in Philosophy*. 5(1): 151–162.

Hájek Alan and Hartman, Stephan (2010). 'Bayesian Epistemology': J. Dancy *et al.* (eds.), *A Companion to Epistemology*. Oxford: Blackwell, 93–106.

Hartmann, Stephan and Sprenger, Jan (2010). 'Bayesian Epistemology' in: S. Bernecker and D. Pritchard (eds.), *Routledge Companion to Epistemology*. London: Routledge 2010, 609–620.

Jehle, David and Fitelson, Branden (2009). 'What is the "Equal Weight View"?', *Episteme* 6(3): 280–293 [doi:10.3366/E1742360009000719]

Joyce, James (2008). 'Bayes' Theorem', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/>.

Kaplan, Mark (1996). *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.

Kelly, Thomas (2005). 'The Epistemic Significance of Disagreement' in Tamar Szabo Gendler/John Hawthorn (eds.), *Oxford Studies in Epistemology* 1. Oxford: Oxford University Press, 167–196.

Kelly, Thomas (2006), 'Evidence' in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/fall2008/entries/evidence/>.

Kelly, Thomas (2010). 'Peer Disagreement and Higher-Order Evidence' in Richard Feldman/Ted A. Warfield (eds.), *Disagreement*. Oxford: Oxford University Press, 111–174.

Kemeny, John G. (1955). 'Fair Bets and Inductive Probabilities', *Journal of Symbolic Logic* 20: 263–273.

Lackey, Jennifer (2005). 'Memory as a generative epistemic source', *Philosophy and Phenomenological Research* 70 (3): 636-658.

Lackey, Jennifer (2010a). 'What Should We Do When We Disagree?' in Tamar Szabo Gendler/John Hawthorn (eds.), *Oxford Studies in Epistemology 3*. Oxford: Oxford University Press, 274–293.

Lackey, Jennifer (2010b). 'A Justificationist View of Disagreement's Epistemic Significance' in A. Haddock, A. Miller/D. Pritchard (eds.), *Social Epistemology*. Oxford: Oxford University Press, 298–325.

Matheson, Jonathan (2009). Conciliatory Views of Disagreement and Higher Order Evidence', *Episteme* 6(3): 269–279 [doi:10.3366/E1742360009000707]

Ramsey, Frank. P (1978). 'Truth and Probability' in Hugh Mellor (ed.): *Foundations: Essays in Philosophy Logic, Mathematics and Economics*. London: Routledge, 58–100 (original article published in 1926).

Ridder, Jeroen de (2013). 'Why Only Externalists Can Be Steadfast', *Erkenntnis*: 1–15.

Sextus Empiricus (2000). 'Outlines of Skepticism'. Edited by Julia Annas and Jonathan Barnes. Cambridge: Cambridge University Press.

Skyrms, Brian (1980). *Causal Necessity*. New Haven: Yale University Press.

Sosa, Ernest (2010). 'The Epistemology of Disagreement', in Adrian Haddock, Alan Millar -and Duncan Pritchard (eds.), *Social Epistemology*. Oxford: Oxford University Press, 278–296.

Talbott, William (2011). 'Bayesian Epistemology', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL =

<http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/>.

Thune, Michael (2010). "Partial Defeaters' and the Epistemology of Disagreement', *The Philosophical Quarterly* 60(239): 355–372.

Van Inwagen, Peter (1996). 'Is It Wrong, Everywhere, Always, and for Anyone to Believe Anything on Insufficient Evidence?', in Jeff Jordan, Daniel Howard-Snyder (eds.), *Faith, Freedom, and Rationality*, Maryland: Rowman & Littlefield, 136–153.

Van Inwagen, Peter (2010). 'We're Right. They're Wrong' in Richard Feldman and Ted A. Warfield (eds.), Disagreement. Oxford: Oxford University Press, 10–28.

Weatherson Brian (2013). 'Disagreement, Philosophical and Otherwise' in David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 54–76.

Jeffrey, Richard C. (1965). *The Logic of Decision*. Chicago and London: University of Chicago Press.

Jaynes, Edwin T. (1968). 'Prior Probabilities' *IEEE Transactions on Systems Science and Cybernetics* 4 (3): 227–241.

Foley, Richard (1992). 'The Epistemology of Beliefs and the Epistemology of Degrees of Belief', *American Philosophical Quarterly* 29 (2): 111–124.

Christensen, David and Lackey, Jennifer (eds.) (2013). *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

Feldman, Richard and Warfield, Ted A. (eds.) (2010. *Disagreement*. Oxford: Oxford University Press

Appendix

Table of Definitions

Table of Cases

Zusammenfassung

Ziel dieser Masterarbeit ist es, mit der Methodik der „Bayesian Epistemology“ (BE) ein neues Licht auf die Debatte der „Epistemology of Disagreement“ (EoD) zu werfen. Begonnen wird dabei damit, die Debatte auf dem Gebiet der EoD selbst kurz vorzustellen. Das Feld an Antwortmöglichkeiten soll aufgezeigt und analysiert werden. Auf zwei Positionen aus diesem Feld wird hernach im speziellen Bezug genommen. Diese zwei Positionen sind der „Equal Weight View“ (EWV), wie er etwa von Christensen vertreten wird, und der von Kelly vorgebrachte „Total Evidence View“ (TEV). Nach der Einführung sollen die entscheidenden Begriffe der „Disagreement“-Debatte im Allgemein und jene der beiden Positionen im Speziellen mit den Mitteln der BE formalisiert werden. Ziel ist es dabei aufzuzeigen, dass was diese beiden Positionen voneinander trennt nicht nur eine unterschiedliche Antwort auf „Disagreement“-Situationen ist, sondern ein unterschiedliches Verständnis von „Higher-Order Evidence“ (HOE). So baut Christensen (2010) eine prinzipielle Unterscheidung zwischen HOE und „Undercutting Defeater“ (UD, dt. unterminierender Anfechtungsgrund) auf. Durch Analyse von HOE und UD mit den Mitteln der BE soll ein besseres Verständnis der Unterschiede von HOE und UD erreicht werden. Ein solches Verständnis ist deshalb von großer Bedeutung, weil diese Dichotomie für den EWV sehr wichtig ist. Christensen argumentiert, dass um HOE in rationaler Weise berücksichtigen zu können, man im Unterschied zu UD Teile seiner ursprünglichen Gründe für eine Überzeugung ignorieren muss, bzw. diese eingeklammert werden müssen. Würde man dies nicht machen, so Christensen, würden die ursprünglichen Gründe immer noch für die Überzeugung sprechen und man müsste die HOE ignorieren, was irrational erscheint. (Christensen 2010: 195) Wichtig ist dies deshalb, weil er auf analoge Weise dafür argumentiert, dass im Fall von „Disagreement“ die ursprünglichen Gründe nicht in die Waagschale geworfen werden dürfen, um die Situation zu beurteilen. (Christensen 2010: 196) Eine genauere Untersuchung der Unterschiede zwischen HOE und UD ist damit also zugleich eine Auseinandersetzung mit wichtigen Argumenten für den EWV. Wenn aber diese Argumentation von Christensen nicht hält, scheint dies zugleich eine Unterstützung für den TEV zu sein. Denn für den TEV spielen sowohl die ursprünglichen Gründe als auch

jene HOE, die durch das „Disagreement“ gegebenen sind, für die Neubeurteilung der doxastischen Einstellung eine wichtige Rolle.

Nachdem damit der EWV auf Basis der formalen Analyse mittels BE kritisiert und für den TEV Partei ergriffen wurde, wendet sich die Arbeit einer Auseinandersetzung mit „Disagreement“-Situationen selbst zu. Mit Hilfe jenes Instrumentariums, das zuvor entwickelt wurde, soll versucht werden eine konkretere Analyse des Einflusses einer „Disagreement“-Situation auf die doxastischen Einstellungen rationaler AkteurInnen zu erstellen. Das Ziel ist dabei der Definition des TEV gerecht zu werden und sowohl die ursprünglichen Gründe der AkteurInnen, als auch die „Disagreement“-Situation selbst, in die Berechnung mit einzubeziehen. Der Vorteil gegenüber dem TEV ist dabei, dass die konkreten Parameter, die einzelne Fälle voneinander unterscheiden, herausgearbeitet werden können, und sich damit von vagen Begriffen, die bloß durch beispielhafte Fälle motiviert werden, zu befreien. Die Ergebnisse dieser Analyse werden am Ende der Arbeit visuell dargestellt werden, um die Auswirkung der Veränderung von relevanten Parametern zu demonstrieren.

Curiculum Vitae

Patrick Lang

Education

10/2011-2/2014(estimate)	**Master of Arts in Philosophy** at the University of Vienna, Vienna, Austria Master Thesis: Bayesian Epistemology of Disagreement (Advisor: Mag. Dr. habil. Sven Bernecker, PhD)
3/2008-5/2011	**Bachelor of Arts in Philosophy** at the University Vienna, Vienna, Austria Bachelor Theses: Kurt Gödel und die Existenz Gottes. Ein Blick auf die Person hinter dem Werk (Advisor: Assoz. Prof. MMag. DDr. Esther Ramharter) Maudlin im Spannungsfeld von Tarski und Kripke. Tarskis Geist als ständiger Begleiter (Advisor: MMag. Dr. Georg Schiemer)
9/2001-6/2007	**Matura** (a-Level) at HTL Pinkafeld, Burgenland, Austria

Professional Experience

9/2013-2/2014	**Student assistant**, University of Vienna, Vienna, Austria Assisting in editorial work for ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth.

3/2013-6/2012	**Student assistant**, University of Vienna, Vienna, Austria Assisting in editorial work for ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth.
10/2012-1/2013	**Student assistant**, University of Vienna, Vienna, Austria Assisting and teaching first year students in introductory courses on theoretical Philosophy held by Mag. Dr. habil. Sven Bernecker, PhD
3/2012-6/2012	**Tutor**, University of Vienna, Vienna, Austria Assisting in the seminar *Ernst Tugendhat zu den anthropologischen Voraussetzungen von Religion und Mystik*, held by ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth
10/2011-2/2012	**Tutor**, University of Vienna, Vienna, Austria Assisting in the seminar *Ernst Cassirer - Zwischen Wissenschafts- und Kulturphilosophieheld*, held by ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth and Donata Romizi, MA
3/2011-7/2011	**Tutor**, University of Vienna, Vienna, Austria Assisting and teaching first year students in introductory courses held by ao. Univ.-Prof. i.R. Dr. Richard Heinrich and ao. Univ.-Prof. i.R. Dr. Herbert Hrachovec

Voluntary Activities

10/2012-Current	Active Member of the *Vienna Forum for Analytical Philosophy* (since 10/2013 as secretary)

Lebenslauf

Patrick Lang

Bildung

10/2011-2/2014(voraussichtlich)	**Master of Arts in Philosophy** an der Universität Wien, Wien, Österreich Masterarbeit: Bayesian Epistemology of Disagreement (Betreuer: Mag. Dr. habil. Sven Bernecker, PhD)
3/2008-5/2011	**Bachelor of Arts in Philosophy** an der Universität Wien, Wien, Österreich Bachelorarbeiten: Kurt Gödel und die Existenz Gottes. Ein Blick auf die Person hinter dem Werk (Betreuerin: Assoz. Prof. MMag. DDr. Esther Ramharter) Maudlin im Spannungsfeld von Tarski und Kripke. Tarskis Geist als ständiger Begleiter (Betreuer: MMag. Dr. Georg Schiemer)
9/2001-6/2007	**Matura** an der HTL Pinkafeld, Burgenland, Österreich

Berufliche Erfahrung

9/2013-2/2014	**Studienassistent**, Universität Wien, Wien, Österreich Assistenz bei Herausgabe eines Sammelbandes für ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth.

3/2013-6/2012	**Studienassistent**, Universität Wien, Wien, Österreich Assistenz bei editorialen Aufgaben für ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth.
10/2012-1/2013	**Studienassistent**, Universität Wien, Wien, Österreich Assistenz und Lehre zur STEOP Vorlesung *Einführung in die theoretische Philosophie*, gehalten von Mag. Dr. habil. Sven Bernecker, PhD
3/2012-6/2012	**Tutor**, Universität Wien, Wien, Österreich Assistenz für das Seminar *Ernst Tugendhat zu den anthropologischen Voraussetzungen von Religion und Mystik*, gehalten von ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth
10/2011-2/2012	**Tutor**, Universität Wien, Wien, Österreich Assistenz für das Seminar *Ernst Cassirer - Zwischen Wissenschafts- und Kulturphilosophieheld*, gehalten von ao. Univ.-Prof. Mag. Dr. Elisabeth Nemeth und Donata Romizi, MA
3/2011-7/2011	**Tutor**, Universität Wien, Wien, Österreich Assistenz zur Ring-Vorlesung *Methoden und Disziplinen der Philosophie*, sowie abhalten von Tutorien für die Einheiten von ao. Univ.-Prof. i.R. Dr. Richard Heinrich und ao. Univ.-Prof. i.R. Dr. Herbert Hrachovec

Zusätzliche Aktivitäten

10/2012-jetzt	Aktives Mitglied im *Wiener Forum für Analytische Philosophie* (seit 10/2013 als Schriftführer)