

RATIONAL DEFERENCE AND EPISTEMIC COMPROMISE:
PRINCIPLES OF REASONING AND UNCERTAINTY

Thesis submitted for the degree of
“Doctor of Philosophy”

By Amir Konigsberg

Submitted to the senate of the Hebrew University of Jerusalem

March 2012

RATIONAL DEFERENCE AND EPISTEMIC COMPROMISE:
PRINCIPLES OF REASONING AND UNCERTAINTY

Thesis submitted for the degree of
“Doctor of Philosophy”

By Amir Konigsberg

Submitted to the senate of the Hebrew University of Jerusalem
March 2012

This work was carried out under the supervision of:
Moshe Halbertal, Eyal Winter, and Gabriel Motzkin

ACKNOWLEDGMENTS

My greatest debts with respect to this thesis are to my family. Anat, Ima, Yitzhak, Yaron, and Tali. Ever since I began my academic studies and throughout the writing of my dissertation I have received endless encouragement and support. I find it difficult to imagine this project without all of you.

Anat, my partner and best friend, has provided me with love, kindness, and endless patience. Thinking through the issues in this dissertation with her and listening to her clear thoughts has made my thinking on the issues pursued in the following pages more mature. My parents, Miki and Yitzhak, have always believed in me and in the projects I pursue, and their influence on me, my writing, and the confidence that I have in my choices is enormous. Yaron, my brother, who was interested in philosophy before I was, always encouraged me to follow this dissertation through, even in hard times, and is always interested in talking philosophy and science. Tali, my sister, challenged me to clarify my ideas, which I also feel made me focus on important problems. Tali also made sure that I never take philosophical endeavors too seriously by making sure to impersonate me when I did.

Throughout the writing of the papers that make up this dissertation I have benefited greatly from comments, suggestions, and conversations with Avishai Margalit, Maya Bar-Hillel, Eyal Winter, Moshe Halbertal, David Enoch, Eldar Shafir, Paul Slovic, Robert Aumann, Adam Elga, Shem Shemy, Eilon Vaadia, Thomas Kelly, Alon Harel, Yuval Eylon, and Edna Ullman-Margalit who passed away when I was in the early stages

of writing my dissertation. Invariably these interactions were enjoyable as well as enlightening.

Others have generously helped in various ways. I would especially like to thank Yael Niv, Hana Shemesh, Yoav Ashkenazi, Roi Zoltan, Ori Weisel, Natalie Meltzer, Eoin Ryan, John Myles White, Romina Goldman, Elinor Harari.

In the course of my studies I have benefited from the support not only of generous individuals but also of generous institutions. The Center for the Study of Rationality at the Hebrew University, the Hebrew University's Presidential Program for Outstanding Doctoral Students, the Tikvah Foundation, and the Max Planck Institute for Science.

At this moment in time I feel that my thinking on the issues dealt with in this dissertation is much clearer than when I began. For me, thinking philosophically tends to also mean making progress in getting clearer about things. And this, at least in part, is what I feel I may have achieved here.

ABSTRACT

This dissertation deals with a number of (what I take to be) philosophically interesting issues that arise from normative considerations involved in the evaluations of beliefs. More specifically, it deals with making up one's mind in uncertain conditions with different kinds of information which is often also limited and ambiguous. The dissertation is composed of three parts. The first part is composed of three papers dealing with the epistemic significance of disagreement and its relation to epistemic compromise. Each of the papers offers a novel contribution to the contemporary debate in the philosophical literature surrounding disagreement. The second part is composed of two papers that, taken broadly, deal with epistemology and aesthetics. Each of the papers raises different normative questions about aesthetic judgments and aesthetic beliefs. Specifically, the first of these papers deals with the possibility of rational aesthetic deference, while the second paper deals with the possibility of non-perceptual aesthetic appreciation. The third and final part of the dissertation includes a single paper which deals with the irrelevance of irrelevant alternatives in choice problems. The paper introduces the notion of decision value and proposes it as an augmented standard to the consequential approach to the evaluation of choice.

The core concern of the papers considered in this dissertation is a form of imperfect or bounded rationality. As opposed to what we might refer to as 'ideal' reasoning, where cognitive, temporal, practical, and perspectival considerations are not typically taken into consideration, in the conception of reasoning I deal with here, the characteristics that differentiate the human reasoner from the ideal reasoner are in fact what I am interested in. Moreover, the type of rationality by which human reasoning can be evaluated is the

type of rationality that will feature throughout the following papers. A large part of the task undertaken in this project is to demarcate exactly when, how, and where these two types of rationalities differ, and when and where each should apply in the evaluation of beliefs and behavior.

The questions addressed here are characteristically about how a person ought to respond to evidence and information seeing as he also knows that his responses are often incorrect. What I have in mind in this project is perhaps a new kind of *prescriptive* epistemology; a discipline that, like ethics, offers practical suggestions for better living. The type of epistemology I envision will perhaps not provide rules of conduct, but it will suggest principles that may aid reasoning by recognizing prevalent fallacious human tendencies. It seems to me that such a proposal is valuable for at least two reasons. Firstly, it relates to reasoning, and we reason all the time, and hence knowing how to do so well is of value. Second, and perhaps more importantly, bad reasoning is far more dangerous and harmful than people recognize, and thus highlighting the kind of errors that bad reasoning can lead to may aid in emphasizing the importance of avoiding it.

The structure of this dissertation is as follows. Chapter 1, “The problem with uniform solutions to peer disagreement,” engages with the contemporary debate surrounding epistemological concerns with disagreement and its relation to epistemic compromise. It challenges the prevalent approach in the literature whereby a uniform solution to peer disagreement is sought for all cases in which disagreements are encountered. The paper argues that different domains appear to call for different responses to disagreement. It argues that the appropriate response to discovering another equally capable person’s opinion on the same matter on the basis of the same evidence, will be different in

different domains. Constructively, the proposal the paper sets forth is, roughly, that the more explicit the procedures for evidential assessment and the more unambiguous the evidence in a particular domain the more reason one has to revise one's belief upon discovering that a peer disagrees. The paper's conclusion is that epistemic compromise can be different in different domains.

Chapter 2, "Compromise, Conciliation, and the Problem of Relevance," offers a new way of considering the basic concerns addressed in the debate involving the epistemic significance of disagreement. The paper aims to highlight what the *real* problem with disagreement is. It is, roughly, the problem of deciding *whether* a revisionary tactic of one's beliefs is appropriate following the discovery of disagreement as well as deciding *which* revisionary tactic is appropriate. The paper shows how this is a slippery and inevitable problem that any discussion of disagreement ought to deal with. The paper proceeds in two stages. In the first stage an outline of the *real* problem of disagreement is set forth. In the second stage three dominant approaches towards disagreement from the literature are examined and the merit of the revisionary responses they propose is assessed. In the course of doing so the paper shows why these approaches and the responses they advance do not address the *real* problem of disagreement. The paper also suggests which of the three approaches prevalent in the literature is most plausible in view of its implicit recognition of the underlying difficulty set forth in the first part of the paper.

In Chapter 3, "Epistemic Value and Epistemic Compromise, a Reply to Moss," I present a criticism of Sarah Moss' recent proposal to use scoring rules as a means of reaching epistemic compromise in disagreements between epistemic peers that have encountered

conflict. The problem I have with Moss' proposal is twofold. Firstly, it appears to involve a double counting of epistemic value. Secondly, it isn't clear whether the notion of epistemic value that Moss appeals to actually involves the type of value that would be acceptable and unproblematic to regard as epistemic.

Chapter 4, "Rational Aesthetic deference," asks whether aesthetic deference ever rational? I argue that it can be. I defend aesthetic deference by means of a distinction between two kinds of aesthetic knowledge, predicative aesthetic knowledge and appreciative aesthetic knowledge. Whereas predicative aesthetic knowledge can be transmitted through deferential belief-formation, it is widely believed that appreciative aesthetic knowledge cannot. Building on this distinction and focusing on the relations between these two kinds of aesthetic knowledge I propose conditions under which aesthetic deference can be rational. Nonetheless, I suggest that there may be non-epistemic norms that undermine deferential aesthetic belief formation.

Chapter 5, "The Acquaintance Principle, aesthetic autonomy, and non-perceptual appreciation," expounds the possibility of what I refer to as non-perceptual aesthetic appreciation, and thus sets for the rational underpinning of aesthetic experiences of conceptual works of art whose appreciation is not grounded in sense perception.

The Acquaintance Principle (AP), roughly, the principle that maintains that it is implausible for a person's aesthetic beliefs to be based on someone else's aesthetic experience, has recently been tied to a debate surrounding the possibility of aesthetic testimony, which, plainly put, deals with the question whether aesthetic knowledge can be acquired through testimony – typically aesthetic and non-aesthetic descriptions communicated from person to person. In this context a number of suggestions have been

put forward opting for a restricted acceptance of AP. This paper is an additional attempt to restrict AP even more than previous suggestions. The paper is composed of two parts. The first part presents and discusses AP, its underlying conceptual foundations, and a number of responses to AP and the view it expresses in the recent literature. The second part argues, by way of demonstration, that AP is varyingly plausible in different domains.

Chapter 6, “Consequential and decision value and the irrelevance of irrelevant alternatives in choice problems,” proposes a distinction between consequential and decision value in the rational assessment of choice. The paper provides an alternative taxonomy for dealing with rational choice problems. It introduces the notion of decision value, and shows that it can explain a wider spectrum of choice problems than can be explained using consequential value alone. Perhaps more importantly, the paper shows that decision value can make sense of choice problems that are considered irrational when assessed only by their consequential value. The paper demonstrates that the consequentialist standard of rational choice, which is regarded as the consensual standard for addressing choice problems, is too strong, as are the internal consistency standards of rationality that are closely related to it.

TABLE OF CONTENTS

ABSTRACT 4

INTRODUCTION12

Part I

THE PROBLEM WITH UNIFORM SOLUTIONS TO PEER DISAGREEMENT.....20

1. INTRODUCTION..... 20

2. THE PROBLEM OF PEER DISAGREEMENT 21

 2.1 *unpacking the problem* 23

 2.2 *disagreement in the literature – from econs to persons*..... 26

 2.3 *being unable to agree to disagree*..... 26

 2.4 *epistemic peers vs. economic agents*..... 30

 2.5 *examples of peer disagreement*..... 36

 2.5.1 *drugs and abortions – two more examples* 43

3. WHAT’S THE PROBLEM WITH A UNIFORM SOLUTION TO PEER DISAGREEMENT? (AND THE PRESENTATION OF MY ARGUMENT). 45

 3.1 *epistemic domains*..... 47

 3.1.1 *How coarsely grained are these domains?* 50

 3.2 *a scale between two extremes* 51

 3.3 *why conciliatory responses are more plausible at the clear and unambiguous end of the epistemic scale* 53

 3.4 *interlude: kornblith and the consensual foundations of conciliation*..... 58

 3.5 *Back to conciliatory responses* 59

 3.6 *when conciliatory responses are less natural*..... 62

4. PRACTICAL VERSUS THEORETICAL SOLUTIONS 65

COMPROMISE, CONCILIATION, AND THE PROBLEM OF RELEVANCE.....69

1. INTRODUCTION 70

2 THE REAL PROBLEM OF DISAGREEMENT – *SETTING THE STAGE*..... 72

 2.1 *Disagreements between ordinary people*..... 73

 2.2 *A brief note on disagreements in the economic literature* 73

 2.3 *Human imperfection and its implications for practical reasoning problems* 74

 2.4 *Imperfection and reasoning* 75

 2.5 *When subjective credence plays against facts of the matter* 76

 2.6 *First-order and second-order evidence* 76

 2.7 *Theoretical and practical prescriptions*..... 79

 2.8 *Practical problems with theoretical prescriptions in disagreement problems* 82

 2.9 *The real problem of disagreement – a practical paradox*..... 84

3 RESPONSES IN THE LITERATURE 85

 3.1 *A case of peer disagreement* 85

 3.2 *The bootstrap approach*..... 86

 3.3 *The conciliatory approach*..... 89

 3.4 *The egalitarian approach* 90

3.5	<i>Back to the conciliatory approach</i>	94
4.	KELLY’S TOTAL EVIDENCE VIEW	94
5.	SO WHAT, AFTER ALL, CAN WE DO?.....	97
EPISTEMIC VALUE AND EPISTEMIC COMPROMISE, A REPLY TO MOSS		99
INTRODUCTION		99
I.....		100
SCORING RULES.....		100
EPISTEMIC VALUE		103
CARING ABOUT CREDENCES.....		104
II		106
WHY THE SCORING RULE SOLUTION MAKE SENSE		106
THE FEASIBILITY OF THE SCORING RULES PROPOSAL		107
DOUBLE-COUNTING		108
EPISTEMIC COMPROMISE AND THE AGENTS TO WHICH IT APPLIES		113
EPISTEMIC VALUE CONSIDERATIONS.....		114

Part II

RATIONAL AESTHETIC DEFERENCE.....		118
Introduction		118
Aesthetic experts		122
Acquaintance		126
Appreciative aesthetic knowledge.....		127
Predicative aesthetic judgments and evaluative verdicts.....		128
Rethinking acquaintance.....		131
Autonomy.....		133
THE ACQUAINTANCE PRINCIPLE, AESTHETIC AUTONOMY, AND AESTHETIC APPRECIATION		137
1.	INTRODUCTION.....	137
2.	THE ACQUAINTANCE PRINCIPLE.....	139
2.1	<i>aesthetic beliefs</i>	141
2.2	<i>weaker versions of the AP</i>	142
2.3	<i>aesthetic autonomy</i>	147
2.4	<i>aesthetic experience and perception</i>	148
2.5	<i>the scope of the aesthetic domain</i>	150
3.	DOMAIN SENSITIVITY	152
3.1	<i>erasing the need to see</i>	152
3.2	<i>non-works of art in the artworld</i>	153
3.3	<i>back to the drawing board</i>	155
3.4	<i>fountain: the aesthetic unimportance of perception</i>	157
3.5	<i>the base of the world and vertical earth kilometer</i>	158
3.6	<i>aesthetic autonomy and aesthetic testimony</i>	160
3.7	<i>can you appreciate a steak without eating it?</i>	161

4. CONCLUDING REMARKS.....	163
----------------------------	-----

Part III

CONSEQUENTIAL AND DECISION VALUE AND THE IRRELEVANCE OF IRRELEVANT ALTERNATIVES IN CHOICE PROBLEMS.....	165
1. INTRODUCTION.....	166
2. CONSEQUENTIAL AND DECISION VALUE.....	167
2.1 <i>a call for a new paradigm</i>	168
2.2 <i>decision value in action</i>	171
3. CONSEQUENTIALISM AND CONSISTENCY.....	173
3.1 <i>a foundational problem with internal consistency</i>	174
4. THE IIA.....	175
5. AMARTYA SEN AND THE EPISTEMIC VALUE OF IRRELEVANT ALTERNATIVES.....	178
6. A DIFFICULTY WITH EPISTEMIC VALUE CONSIDERATIONS.....	180
6.1 <i>where does this leave us?</i>	182
6.2 <i>epistemic value and decision value</i>	183
7. THE EPISTEMIC RELEVANCE OF CONSEQUENTIALLY IRRELEVANT ALTERNATIVES.....	184
7.1 <i>a case of decision value</i>	185
8. IS DECISION VALUE NOT SIMPLY A CONCLUSIVE FORM OF CONSEQUENTIAL VALUE?.....	186
9. CONCLUSION.....	187
BIBLIOGRAPHY.....	188
תקציר בעברית.....	199

INTRODUCTION

Beliefs, like decisions, art, and relationships, can be normatively evaluated. Just as a person might regard an act as wrong, a painting as kitsch, and a relationship as good, they may regard a belief as implausible, improper, warranted, unwarranted, certain or uncertain. Beliefs are, in this sense, the kinds of things for which critical appraisal is appropriate. Needless to say that the critical appraisal of beliefs is a hugely important means of ensuring that proper beliefs are held and that other beliefs and actions based on them are properly grounded.

In the following pages, I pursue a number of (what I take to be) philosophically interesting issues that arise from normative considerations involved in the critical appraisal of beliefs. I think this is an interesting, important and complicated subject and my interest in it has concerned me for longer than I have been a student of philosophy. Perhaps more importantly, I believe that contributions in this area can be practically useful in prescribing measures and principles for reasoning.

“Man is a credulous animal, and must believe something; in the absence of good ground for belief, he will be satisfied with bad ones” (Russell 1996, 111). This observation was made by Bertrand Russell and I feel that it expresses my point of departure in the present intellectual exploration. The statement can be read as calling for a corrective examination of the standards for the evaluation of beliefs. In this spirit, the focus of the following inquiry is human reasoning observed in the wider context of human fallibility.

Human reasoners are not flawless gatherers of information. Neither are they faultless at interpreting information and inferring conclusions from it. Time restrictions, cognitive limitations, as well as biases and other varieties of perceptual, cognitive and emotional distortions contribute to the error-prone reasoning that is characteristic of humans. This often leads to unexpected and erroneous responses to evidence. The exploratory approach exhibited in the following papers recognizes that human reasoners routinely acknowledge their own imperfections (often resulting from the above-mentioned causes) and devise strategies to deal with it.

This inquiry is epistemological, broadly construed. But it is concerned with a distinctive kind of epistemology that I believe is suited to a human reasoner's 'belief management' considered in the wider context of his own fallibility. Edna Ullmann-Margalit introduced me to the notion of belief management. The notion can, roughly, be taken to describe the 'managerial' responsibility with which the human reasoner is tasked with in regulating his beliefs in light of the changing states of affairs around him. The concept is especially useful because it emphasizes that ongoing critical appraisal – the updating, modifying, and revising of beliefs - is a human responsibility. A significant portion of this study seeks to bring to light some the principles by which beliefs are 'managed.'

Philosophical problems, Wittgenstein once said, are philosopher's problems. What he seems to have meant by this is that often, the problems which philosophers engage with are fashioned by their own confusions. I think that this may well be true of some philosophical concerns, but I do not think this is always so. Indeed, a number of the philosophical concerns that I address in the present collection of papers do not seem to me to only be the concerns of the philosopher. Instead, I regard them as *real* concerns,

which the philosopher may bring to light but that he is in no way responsible for creating. What I have in mind here are concerns commonly encountered in practical reasoning and deliberation, concerns that responsible individuals in all walks of life often find themselves engaging with, if not theoretically, then practically.

The papers included here were written over the course of several years while I was a doctoral student at The Center for the Study of Rationality and Interactive Decision Theory at the Hebrew University in Jerusalem and at the Psychology Department of Princeton University.

The interdisciplinary nature of the Center for the Study of Rationality and the mixed disciplinary affiliation of its members have significantly influenced my intellectual growth, the subjects addressed in the papers presented here, and the methodological toolkit used to approach these subjects. The welcoming environment of the Center, the accessibility of its members, and the array of perspectives that these surroundings afforded my interest in questions of rationality and decision making, were invaluable. So too, the setting at Princeton's Department of Psychology, my affiliation with a lab dealing with conditional learning and experimental decision making, as well as the influence of economists and researchers in social and cognitive psychology have given me a broader and enhanced perspective with regard to the questions perused here. I feel extremely fortunate for having been part of these fruitful and welcoming environments.

While a significant portion of the issues with which I deal in the following chapters fall within the province of epistemology broadly construed, the conclusions that I arrive at have implications for other areas of philosophy as well as for closely related fields such

as rationality, decision theory and reasoning. This is perhaps the correct place to note that this project has been influenced by advances in the recent decades in these fields. Indeed, I feel that the significance of these advances for philosophy in general and epistemology in particular has, to a great extent, been under-appreciated. In particular I have in mind the core principles of the Judgment and Decision Making approach in social and cognitive psychology, the psychology of Risk Perception, and Behavioral Economics. Themes, concepts and conclusions from these adjacent disciplines feature throughout the present inquiry.

Although I hope to have made some contribution to understanding the issues I have chosen to deal with here, I would be disappointed if the interest and value of the present work was exhausted by whatever bearing it does have on these questions. It is my hope that various ideas that I discuss along the way might prove interesting and fruitful even when considered apart from the larger intellectual frameworks in which they come embedded in the present study. In particular, I am thinking of the idea of epistemic compromise; the emphasis on the epistemic significance of disagreements and their corrective role; the distinction between first-order and second-order evidence in the assessment of beliefs; the role of subjective judgment and its epistemic relation to independent information; the prescriptive conception of epistemology which I am proposing; the idea of rational deference and its plausibility in various epistemic and non-epistemic domains; the significance of prior probabilities for diagnosticity; the relation between rationality and consistency; and the epistemic significance of testimony.

While the underlying theme of the papers presented here relates to normative considerations in the critical appraisal of beliefs, each of the papers is self-standing and

can be read independently of the others. The format I have chosen for this dissertation is a compilation of seven independent papers, presented as chapters. While the themes of these papers occasionally overlap, each of the papers stands on its own and should be read as such.

The structure of this dissertation is as follows. Chapter 1, “The problem with uniform solutions to peer disagreement,” engages with the contemporary debate surrounding epistemological concerns with disagreement. It challenges the prevalent approach in the literature whereby a uniform solution to peer disagreement is sought for all cases in which disagreements are encountered. The paper argues that different domains appear to call for different responses to disagreement. It argues that the appropriate response to discovering another equally capable person’s opinion on the same matter on the basis of the same evidence, will be different in different domains. Constructively, the proposal the paper sets forth is, roughly, that the more explicit the procedures for evidential assessment and the more unambiguous the evidence in a particular domain the more reason one has to revise one’s belief upon discovering that a peer disagrees.

Chapter 2, “Compromise, Conciliation, and the Problem of Relevance,” offers a new way of considering the basic concerns addressed in the debate involving the epistemic significance of disagreement. The paper aims to highlight what the real problem with disagreement *is*. It is, roughly, the problem of deciding *whether* a revisionary tactic is appropriate following the discovery of disagreement as well as deciding *which* revisionary tactic is appropriate. The paper shows how this is a slippery and inevitable problem that any discussion of disagreement ought to deal with. The paper proceeds in two stages. In the first stage an outline of the real problem of disagreement is set forth. In

the second stage three dominant approaches towards disagreement from the literature are examined and the merit of the revisionary responses they propose is examined. In the course of doing so the paper shows why these approaches and the responses they advance do not address the *real* problem of disagreement. The paper also suggests which of the three approaches prevalent in the literature is most plausible in view of its implicit recognition of the underlying difficulty set forth in the first part of the paper.

In Chapter 3, “Epistemic Value and Epistemic Compromise, a Reply to Moss,” I present a criticism of Sarah Moss’ recent proposal to use scoring rules as a means of reaching epistemic compromise in disagreements between epistemic peers that have encountered conflict. The problem I have with Moss’ proposal is twofold. Firstly, it appears to involve a double counting of epistemic value. Secondly, it isn’t clear whether the notion of epistemic value that Moss appeals to actually involves the type of value that would be acceptable and unproblematic to regard as epistemic.

Chapter 4, “Rational Aesthetic Deference,” asks whether aesthetic deference is ever rational. I argue that it can be. I defend aesthetic deference by means of a distinction between two kinds of aesthetic knowledge, predicative aesthetic knowledge and appreciative aesthetic knowledge. Whereas predicative aesthetic knowledge can be transmitted through deferential belief-formation, it is widely believed that appreciative aesthetic knowledge cannot. Building on this distinction and focusing on the relations between these two kinds of aesthetic knowledge I propose conditions under which aesthetic deference can be rational. Nonetheless, I suggest that there may be non-epistemic norms that undermine deferential aesthetic belief formation.

Chapter 5, “The Acquaintance Principle, aesthetic autonomy, and non-perceptual appreciation,” expounds the possibility of what is referred to as non-perceptual aesthetic appreciation, and thus sets for the rational underpinning of aesthetic experiences of conceptual works of art whose appreciation is not grounded in sense perception.

The Acquaintance Principle (AP), roughly, the principle that maintains that it is implausible for a person’s aesthetic beliefs to be based on someone else’s aesthetic experience, has recently been tied to a debate surrounding the possibility of aesthetic testimony, which, plainly put, deals with the question whether aesthetic knowledge can be acquired through testimony – typically aesthetic and non-aesthetic descriptions communicated from person to person. In this context a number of suggestions have been put forward opting for a restricted acceptance of AP. This paper is an additional attempt to restrict AP even more than previous suggestions. The paper is composed of two parts. The first part presents and discusses AP, its underlying conceptual foundations, and a number of responses to AP and the view it expresses in the recent literature. The second part argues, by way of demonstration, that AP is varyingly plausible in different domains.

Chapter 6, “Consequential and decision value and the irrelevance of irrelevant alternatives in choice problems,” proposes a distinction between consequential and decision value in the rational assessment of choice. The paper provides an alternative taxonomy for dealing with rational choice problems. It introduces the notion of decision value, and shows that it can explain a wider spectrum of choice problems than can be explained using consequential value alone. Perhaps more importantly, the paper shows

that decision value can make sense of choice problems that are considered irrational when assessed only by their consequential value. The paper demonstrates that the consequentialist standard of rational choice, which is regarded as the consensual standard for addressing choice problems, is too strong, as are the internal consistency standards of rationality that are closely related to it.

THE PROBLEM WITH UNIFORM SOLUTIONS TO PEER DISAGREEMENT

Abstract: in this paper I challenge a prevailing approach in the philosophical literature according to which there exist uniform solutions to disagreement problems. I argue that the appropriate response to disagreement is different in different domains. Roughly speaking, the more explicit the procedures for evidential assessment and the more unambiguous the evidence in a particular domain the more reason one has to revise one's belief upon discovering that a peer disagrees.

1. INTRODUCTION¹

Contributors to the recent debate surrounding disagreement have sought to provide a uniform response to cases in which epistemic peers disagree about the epistemic import of a shared body of evidence, no matter what kind of evidence they are disagreeing about. The varied cases addressed in the literature have included examples of disagreement about restaurant bills (Christensen 2007), court verdicts (Rosen 2001), weather forecasting (Kelly 2009), chess (Kelly 2005), morality (Elga 2007; Enoch 2009), religious beliefs (Feldman 2007), and even disagreements about philosophical disagreements (van Inwagen 1996; Enoch 2010; Kornblith 2010). The equal treatment of these varied cases of peer disagreement has motivated the search for a uniform response to peer disagreement wherever it is encountered.

¹ For a published version of this chapter in *Theoria* see: (Königsberg *forthcoming*).

In this paper I challenge this prevalent approach in the literature. I grant the notion of epistemic peer and accept that being a peer may amount to the same thing in different domains; nonetheless I contend that different domains appear to call for different responses to disagreement. I will argue that the appropriate response to finding out about a disagreement with a peer is different in different domains. Roughly speaking, the more explicit the procedures for evidential assessment and the more unambiguous the evidence in a particular domain the more reason one has to revise one's belief upon discovering that a peer disagrees.

The paper is composed of three sections. I begin by outlining the problem of peer disagreement and point to some examples characteristic of how it has been addressed in the literature, in a way that exhibits the uniform treatment I want to undermine. I then present an argument for why I think the uniform treatment of disagreement in the literature is wrong and why an unvarying response to peer disagreement is implausible. In the final section I respond to an initially powerful objection to my argument.

2. THE PROBLEM OF PEER DISAGREEMENT

I shall begin by trying to capture the problem of peer disagreement as it has been presented in the literature. I take the subsequent presentation of the problem to be in the spirit of the contemporary debate. I will set up my discussion of the problem fairly simplistically; I will then add complications that may seem applicable in some cases of disagreement. But these will be stripped away to an extent, to arrive at a version of the problem which we can see to be the essence of the problem that is responsible to ordinary

life encounters with disagreement. Here we shall attend to the philosophical nub at the centre, with extra psychological and other considerations cleared away.

In the sense I use here, which follows that of the literature, two people are epistemic peers in some domain ψ , if and only if they are equally likely to be correct when giving an answer to an unspecified question about ψ ². In other words, they are epistemic peers in ψ if and only if they have the same probability of answering a question correctly in ψ . I am leaving the nature of ψ unspecified for the moment, to be intuitively understood, while I focus on bringing forth the central notion of disagreement³.

Now, a question is drawn at random from an imaginary box containing all the questions pertaining to ψ . Each of the peers answers it. Then each discovers that his peer gave a different answer: A answers *not-p* and discovers that B answered *p*. And similarly, B answers *p* and discovers that A answered *not-p*, to take a pseudo-particularized description of *a* case of disagreement.

² I take this formulation to be in the spirit of how the problem has been addressed in the literature. See: (Christensen 2007; Elga 2007a; Feldman 2006; Feldman and Warfield 2010; Kelly 2005; Kelly 2009; Enoch 2010).

³ Intuitively at least, there appear to be a variety of ways by which peerhood can be established. But at present I find it sufficient to note that there appear to be a continuum of possible notions of epistemic peerhood, ranging from ‘mere’ equal reliability in a given domain (e.g., (Christensen 2007; Feldman 2006; Enoch 2010) on the one end of the spectrum, to equal familiarity with the evidence combined with broad agreement about epistemic values and equal awareness of the limitations of one’s own knowledge, on the other ((Gelfert 2011), which also contains a survey of existing notions of epistemic peerhood).

Now assuming that neither A nor B have any reason to believe that either of them is functioning epistemically better or worse than usual with respect to ψ , how then should A update his belief about p seeing as he now knows what B believes about it, assuming that he also knows that B is his epistemic peer in matters that relate to ψ ? This, in a nutshell, is the problem of peer disagreement.

2.1 UNPACKING THE PROBLEM

Finding out that an epistemic peer both draws conclusions from the same evidence as you but that these are different conclusions raises immediate questions about what ought to be your appropriate response. The most pressing question seems to be whether you ought to change something about your present belief after finding out what your peer believes about the same matter. The question is especially perplexing since before this problem arose, as peers, the two of you were equally likely to be right (we assume that you are both aware of this). And, given that nothing in the evidential circumstances pertaining to ψ or to any question relating to ψ has actually changed, other than discovering that you disagree about what the evidence *means*, it would seem that neither of you, now that you are aware of the disagreement, has any non-question-begging reason to think yourself right⁴.

Now, if we take someone's competence in a particular domain to reflect their chances of being successful in responding to a question about that domain, then because A and B are equally competent in ψ , however sure A may be that *not-p* is the case he must also

⁴ A related question is whether the disagreement itself gives you an epistemic reason to modify your belief in any way, and if so, in what way. Enoch deals with this question (Enoch 2010, 6).

assume that B is just as sure that p is the case, and that B is just as likely to be right. Indeed, because they are equally likely to be right, in responding to the encounter with B's surprising belief, A cannot justifiably grant his own belief that $not-p$ more weight than he grants B's belief that p . And perhaps, if A has reasons for believing as he does, however compelling he may find his reasons for believing $not-p$ to be, he must also assume that B feels the same way about his own reasons for believing p .

Now, for A to give his own belief more weight than B's would be for him to override his own knowledge that B is a peer. Because A and B are both just as likely to be right, it seems unlawful for A to leverage the force of the reasons he has for believing $not-p$ and grant them extra weight in deliberating how to respond to B's belief. To do so would be for A to ignore his prior knowledge that whatever reasons he may have, B's belief is just as likely to be right. For A to leverage the preference he has for his own belief would be for him to grant it an illegitimate epistemic privilege. In effect this would be like his using the belief that he is trying to prove to support the belief that he is trying to prove.⁵ The evident circularity of this brings out again why it is illegitimate for A to neglect the fact that B is just as competent as he is and to proceed to dismiss B's present belief that p on the basis of the conviction that he himself has about $not-p$.

In cases where the evidence in fact allows a decision to be made about who is right, A or B, that A is right suggests that B's status as an epistemic peer to A be degraded, for B has answered a question about ψ wrongly that he could have and should have answered correctly, while A has answered it correctly. However, in this case there is no evident

⁵ Compare: (Cohen 2002).

way to decide between A and B. Thus, and for similar reasons to those educed in the previous paragraph, for A to demote B from the status of peer only on the basis of their present disagreement and for A then to take this demotion as a *reason* for him to think himself right is also impermissible. While it may seem legitimate for A to no longer regard B as his peer because they presently disagree, taking their disagreement to support his own belief would be for A to neglect the base-rate information about their prior equal competence. Given the mutually recognized prior probabilities for both A and B's success in ψ , nothing about the present evidential situation seems to justify deviating from what the prior probabilities warrant, and thus neither A nor B are justified in thinking themselves better placed in relation to p . Just as the disagreement could mean that B is wrong, it could also mean that A is wrong.

Therefore in light of A and B's previously established equal epistemic competence in ψ , the rule would seem to be that the epistemic force of any consideration which A possesses in favour of believing p should be equally attributed to B's considerations for believing *not-p*, whether or not B's considerations seem convincing to A. And likewise for B; any consideration that B may feel warrants his thinking himself right on this occasion must be assumed to have a correlate in A's considerations for believing otherwise. This leaves both A and B in a symmetrical situation with regard to the status they ought to grant to each other's beliefs. But all the same, while the base-rate information makes the situation symmetrical, these base-rate considerations may stand in tension with what A or B believe.

2.2 DISAGREEMENT IN THE LITERATURE – FROM ECONS TO PERSONS

In a different context and under a different name, peer disagreements were initially discussed in economic theory, rather than in philosophy. But the peers and the disagreements referred to in these earlier discussions were, on the whole, abstract idealizations of encounters, not between real people, but rather between economically rational agents, “Econs,” as the neo-classical economical conceptions of rational man have recently been referred to (Thaler and Sunstein 2008).⁶ The idealized agents whose epistemic performance was discussed in these early accounts (henceforth I will relate to these as *Econs*) are allegedly supposed to have been able to “think like Albert Einstein, store as much memory as IBM’s Big Blue, and exercise the willpower of Mahatma Ghandi” (Thaler and Sunstein 2008, 6). These agents were regarded as ideals of cognitive behavior rather than of real *human* cognitive behavior. And so too, the epistemic similarities between these agents, what we might take to be equivalent to their ‘peerhood,’ and the suggested disagreements between them, were conceived of quite differently to the way they were understood in later philosophical accounts. For the purpose of my main argument it will be helpful to briefly point out the main differences between these two debates surrounding disagreement.

2.3 BEING UNABLE TO AGREE TO DISAGREE

Robert Aumann’s seminal *Agreeing to Disagree* (Aumann 1976) provided a starting point for debating an interesting and especially surprising result that gave rise to a wide

⁶ *Homo economicus* is the term used by most critical accounts of the neo-classical conception of rational man.

area of research within economic theory. Aumann's result was this: mutually respectful, honest, and rational debaters cannot disagree on any factual matter once they know each other's opinions. Aumann's outcome is surprising given how ordinary people who are at least mostly rational and generally competent go about their belief-forming business.

While Aumann's proof is built on a possible-world semantics formalism so powerful that he apologizes⁷, the theorem he proposes can be put roughly as follows:

- If two agents have the same prior probabilities and the posterior evaluation that they assign to some proposition p is common knowledge, i.e., transparent and mutually accessible to each of them, then those posteriors must – eventually – be equal. Moreover, even if the posteriors were initially derived from different bodies of evidence they will gradually converge to a common equilibrium posterior.⁸

The heart of Aumann's "disagreement theorem" is that rational agents with the same priors cannot agree to disagree since their common knowledge of each other's responses will gradually, through an iterative process, lead both to update their posteriors (in accordance with one another's posteriors). And this is a process that will inevitably lead to consensus either that p is the case or not the case.

⁷ "We publish this note with some diffidence, since once one has the appropriate framework, it is mathematically trivial" (Aumann 1976).

⁸ On a related matter, Geanakoplos's and Polemarchakis's paper "We can't disagree forever," (John Geanakoplos and Heracles M. Polemarchakis 1982) discusses the sequence of events as two rational debaters come to agreement.

From the perspective of ordinary human agents this is an unrealistic result. It is unrealistic because there seem to be many cases where two or more people *can* hold different, at times opposing, beliefs, with both being aware of this, and without either one being considered blatantly irrational in doing so. What's more, it isn't obvious that consensus is always requisite for concluding a rational debate between human agents.

In non-factual domains it may seem natural that consensus needn't be reached simply because there is no fact-of-the-matter to agree about. And this is what seems to make disagreement and the toleration of disagreement more plausible in non-factual debates. Yet all the same, it isn't immediately clear that consensus is always required in factual domains where there *is* a fact-of-the-matter, or that disagreeing on factual matters must always entail that one of the disagreeing parties is being irrational. This seems to be because, as Gideon Rosen rightly observes: "it would appear to be a fact of epistemic life that a careful review of the evidence does not guarantee consensus, even among thoughtful and otherwise rational investigators" (Rosen 2001, 71–72). 'Epistemic life,' if I understand Rosen correctly, refers to the type of circumstances that ordinary people encounter, and thus it is the inherent imperfection and consequent fallibility that even the rational and thoughtful among us are prone to that entails that reviewing the evidence will not always lead to agreement⁹.

⁹ A realist will surely hold that there are facts of the matter that no one need be able to decide on as they are beyond our epistemic reach. I suppose such domains are those where rational debaters will eventually reach a consensus of withholding judgment, but more ordinary people might operate on a hunch. But then the hunch is either somewhat unreal – plays out to have no consequences in the world – or provokes discovery

Factual questions such as what killed the dinosaurs or when the Neanderthals became extinct are normally regarded as questions that in principal can only have a single true answer. And in addition to their particular justification or plausibility, it is also by their relative proximity to this supposed single true answer that varying responses to such questions are evaluated¹⁰. But as is often the case in ordinary epistemic life, there is no one belief that is regarded as correct given the evidence. And in such cases it is precisely because a review of the evidence by equally competent people does not always lead to agreement that the question about whom or what is right arises. Moreover, because neither party in the disagreement has any incremental information about who is right or who is wrong, the normative question is what the appropriate response ought to be¹¹.

Yet according to Aumann's disagreement theorem and the past thirty or so years of research that stemmed from it, strictly rational agents – Econs, that is - cannot agree to disagree.¹² This result stems from the fact that Econs are what we might refer to as *economically rational agents*, agents that update their beliefs consistently and predictably

of further evidence showing the hunch is true/false (as in the bastard verdict of Scots law, neither guilty nor innocent, but “not proven”). I am grateful to Eoin Ryan for this point.

¹⁰ Richard Feldman (Feldman 2007) has proposed what he calls “The Uniqueness Thesis” according to which “a body of evidence justifies at most one proposition out of a competing set of propositions (e.g., one theory out of a bunch of exclusive alternatives) and that it justifies at most one attitude toward any particular proposition.”

¹¹ It is important to point out that if we were rational Econs, and insufficient evidence exists for decisions about dinosaurs, then we would all contentedly abstain from decision.

¹² For a review of the literature that followed the publication of Aumann's agreement theorem, see: (Cowen and Hanson 2004) and the bibliography at the end of: (Giacomo Bonanno and Klaus Nehring 1997).

upon encounter with new information. These are not imperfect rational agents like you and I, and the majority of the people that we know whose import from evidence is not always consistent or predictable. But it is these latter disagreements of a more familiar kind, such as those that normal, imperfectly rational people habitually encounter, that philosophers have recently begun to discuss.¹³ To compare the consequences of the economists' discussion of disagreement to the philosophical discussions of disagreement is problematic on account of the disparate nature of the agents being referred to in each of these debates.

The differences between the economic and the philosophical concerns with disagreement can be outlined in relation to three factors: 1) peerhood; 2) disagreement, and 3) context. These differences are worth noting so as to better specify the normative question faced by disagreeing peers as it shall be attended to here. But, additionally, the economic conception of agency that underlies the economic discussions of disagreement seems to have found expression within the philosophical discussions of disagreement and may also be a reason for the uniform approach to disagreement that I will be arguing against. Hence, it is also in light of the present aim of showing that a uniform approach to peer disagreement is wrong, that it will be useful to call attention to the following distinctions.

2.4 EPISTEMIC PEERS VS. ECONOMIC AGENTS

I) *Peerhood*: as Thomas Kelly has rightly pointed out (Kelly 2005), while epistemic peers may be equally likely to be right in a particular area, their being equally likely to be right does not ensure that they will import *the same* beliefs from the same evidence. Contrary

¹³ Kelly addresses this matter at length in (Kelly 2005, sec. 3).

to Econs, epistemic peers do not have the epistemic counterparts of common priors and consistent conditionalization, which, if they did have, would seem to ensure that they would respond the same to the same evidence. Yet what epistemic peers do have is an equal likelihood of being right. But nothing about this ensures that they will respond the same to a given epistemic question. Moreover, because even bad epistemic agents can be peers – so long as they are *as* bad as each other – then it becomes clear that epistemic peers can respond imperfectly. This means that even if they are equally likely to be right in their beliefs, their beliefs do not have to converge.

The common prior assumption, which until recently was an article of faith among economists, says that differences in beliefs between agents can be completely explained by differences in information.¹⁴ The idea is that agents start off with the same prior beliefs and thereafter update according to the information they receive. If there is later a difference in beliefs between agents, it is due to the fact that they have in the meantime received different information.¹⁵ But what in fact characterizes the philosophical problem of peer disagreement is that people – indeed, rational people - do appeal to the same body of evidence; that is, they *do* receive the same information, and yet they *still* disagree.¹⁶

¹⁴ Hanson (Hanson 2003) suggests that disagreements between “Bayesian’s with computational limitations” are not fundamentally due to differing information.

¹⁵ For an overview see: (Halpern 2000, 133–134).

¹⁶ See also: (Kelly 2005, sec. 3).

And it is what each person ought to take this situation to mean that raises normative questions about what it *ought* to mean¹⁷.

The philosophical problem of peer disagreement relates to human fallibility considered in the wider context of human imperfection. And it assumes that it is a fact of epistemic life that different people are not guaranteed to draw the same conclusions from a shared body of evidence. Moreover, it is because of this familiar situation that the peers who disagree will need to figure out for themselves whether it is they that are wrong about the evidence.

While the philosophical problem of disagreement considers human fallibility, it requires a normative solution. In fact it is actually because it takes into account the imperfect circumstances in which people operate - the at times ambiguous evidence that people encounter and their at times imperfect responses to this evidence - that the normative question about the appropriate response to disagreement arises.

Philosophical questions regarding the normative significance of peer disagreements consider how one ought to proceed after finding out that an esteemed, honest, and no less competent person is discovered to disagree about a shared body of evidence. While the peers referred to do not share common priors, they do share what we might refer to as *common epistemic credentials*, credentials that relate to their past achievements and their

¹⁷ It is worth noting here that Enoch (Enoch 2010) has given a more sustained defense of the significance of the first-person perspective in determining how we should respond to disagreement.

capabilities, in light of which they have an equal probability of getting things right. It is for the most part because of these credentials that they are regarded as epistemic peers¹⁸.

II) *Disagreement*: the way we define what an epistemic peer is will directly impact the nature of our concern with peer disagreements. To see this, consider a disagreement between two Econs. Because of how we conceive them, both Econs are expected to respond *the same* to a particular body of evidence. And therefore given an equally accessible shared body of evidence, either both Econs respond the same to the evidence and do not disagree, or else they do disagree because either or both of them are not responding rationally. It is on the foundations of scenarios such as this that “no bet” theorems in economics have shown that rational agents will not participate in betting markets. Because in such cases the very fact that someone is willing to accept your bet is regarded as evidence that you are wrong.¹⁹

Yet if we were to instead understand these agents not as Econs but as epistemic peers as we have defined them, i.e., as human agents who are fallible and thus not certain to issue the same response to an unspecified question in ψ , then what a disagreement between

¹⁸ I say “for the most part” because people may regard each other as peers for reasons other than their similar credentials and capabilities. This is not to say that they will be justified in doing so, but rather it merely states that it is a possibility. David Enoch makes an important distinction relating to this and is worth quoting fully here: “We can ask, first, how to revise our beliefs in the face of disagreement with someone who is *in fact* our peer (that is, someone who is in fact equally likely as we are to get things right here). Or second, we can ask how to revise our beliefs in the face of disagreement with someone whom *we take to be* our peer. Or third, we can ask how to revise our beliefs in the face of disagreement with someone whom we *justifiably take to be* our peer” (Enoch 2010, sec. 5).

¹⁹ See the classic paper on this: (Milgrom and Stokey 1982) and recent work by (Cowen and Hansen 2004).

peers may be taken to mean is different. While divergence in opinion for Econs implies irrationality, divergence between epistemic peers needn't mean that either agent has diverted from the likelihoods of their expected performance²⁰. And this is what leads us to ask what the appropriate response is to the discovery of disagreement.

III) *Context*: ordinarily, if two people who are peers in some area find themselves in disagreement about Ω , we would consider it reasonable for the both of them to re-evaluate what they think about Ω . We would regard this as the reasonable thing to do in part because we typically recognize our own fallible nature, and so we know that it is common for us, and for beings like us, to make mistakes and that often enough we are wrong and others are right. And for this reason it appears that after finding out that a person that is no less competent and honest than you disagrees about Ω , rechecking your belief (if there is no cost incurred in doing so) by reevaluating the evidence is a common sense, smart, and sensible thing to do.²¹ But rechecking that what you believe is correct is not always an option. And even if it is an option, re-evaluating evidence will not necessarily grant additional epistemic support to your belief, since re-confirming what you already believe isn't entirely the same thing as gaining incremental support for your belief.

²⁰ Consider a situation in which A and B are epistemic peers – they are both 0.8 likely to be right on any question in ψ . Because they are imperfect – i.e., their probability of being right is < 1 , then when they disagree it is still reasonable for each of them to believe that they may be right and the other wrong.

²¹ This is also true if a person is less competent than you; you might be mistaken (because you're not perfect) and he might happen to be right. The same is true for a person less honest than you. Assuming he is not dishonest all of the time, he may well be being honest *this* time.

Positive feedback loops about psychological affirmation are a common phenomenon and often a person will want to maintain a particular belief because modifying it will have an emotional cost²². In such cases, so as to mitigate the expected emotional cost, the evidence that is encountered is taken to be supportive of the beliefs that one already possesses, thus maintaining the status quo. Moreover, often, encountering doubt may make an individual check his beliefs only to reinforce them (groundlessly) to himself through a mechanism of psychological self-affirmation – so that encountering doubt may make a person *more* attached to his beliefs rather than less²³. My point in mentioning this is that the philosophical problem of disagreement is essentially a theoretical problem, not a practical problem. And what I mean by this is that the central question about how you *should* respond following the discovery of an opposing belief held by your peer does not leave open the possibility of re-checking the evidence and re-evaluating your considerations. Instead it assumes that after you have formed your own opinion you discover that a peer holds an opposing view. And the problem debated is how you should update your belief following this discovery. More specifically it is the epistemic significance of the disagreement itself that is questioned, and whether it gives you a *reason* to change your belief in some way, and if it does, then in what way. There is no talk of revisiting the evidence to re-affirm your belief. I will therefore assume that the problem of disagreement relates to how one should update one's belief following the

²² A classic paper on the topic and a cornerstone in the field of social cognition is Ziva Kunda's *The Case for Motivated Reasoning* (Kunda 1990).

²³ On the related phenomenon of belief (and attitude) polarization see: (Lord, Ross, and Lepper 1979) as well as Kelly's excellent discussion in: (Kelly 2008). For a more general overview of biased distortions of evidence and psychological self-affirmation see: (Gilovich 1993, pt. 2).

discovery that a peer holds an opposing view. I will not consider the possibility of re-affirming one's belief by reconsidering the evidence. As we have seen, doing so may not grant additional epistemic support, nor is it a possibility that can be considered in light of the theoretical nature of the problem being discussed and how it has been set up in the literature.

These distinct conceptions of (I) peerhood, (II) disagreement, and (III) context, that characterize the philosophical problem of disagreement between epistemic peers, locate the normative problem of disagreement in the context of uncertain circumstances encountered by imperfect agents whose responses to evidence and to disagreement are not infallible. And it is within this context that the question of what ought to be the appropriate response is addressed.

2.5 EXAMPLES OF PEER DISAGREEMENT

Let us turn now to some examples of peer disagreement which have been discussed in the philosophical literature. Note that each of these examples is supposed to capture different types of cases in which agents with an equal probability of getting things right arrive at incompatible conclusions. In each of the examples it is the way in which the agents ought to respond that is debated.

In the upcoming pages there are two factors that will be taken into consideration in assessing responses to each of the following examples of peer disagreement: 1) the symmetric treatment of the disagreeers and 2) the personal convictions that each of the disagreeers has about their beliefs. In the first two examples – David Christensen's restaurant bill and Peter van Inwagen's disagreement with David Lewis – we are lead to

the Equal Weight View (Elga 2007a) and the Split the Difference View (Christensen 2007), two views which are at root the same in terms of treating the disagreeers symmetrically. Under these views, we have the elimination of the relevance of personal conviction. In the third and fourth examples – the drug example and the abortion example – conciliatory responses such as the Equal Weight View and the Split the Difference View seem to incur a higher cost and thus appear less plausible. In these cases personal conviction does not only play a justificatory role, but rather appears to be constitutive of the debated beliefs themselves.

Let us start off by considering the following paradigmatic example of peer disagreement taken from Christensen (Christensen 2011) and adapted from (Christensen 2007):

After a nice restaurant meal, my friend and I decide to tip 20% and split the check, rounding up to the nearest dollar. As we have done many times, we do the math in our heads. We have long and equally good track records at this (in the cases where we've disagreed, checking with a calculator has shown us right equally frequently); and I have no reason (such as those involving alertness or tiredness, or differential consumption of coffee or wine) for suspecting one of us to be especially good, or bad, at the current reasoning task. I come up with \$43; but then my friend announces that she got \$45.

The dominant view with regard to resolving cases such as that described in the restaurant bill example is to consider the symmetrical nature of the disagreeing parties. Thus the person who arrived at \$43 should, following his friend's announcement, be less confident

in his own result than he had been before finding out about the disagreement, and should be more confident than he had been that the result is \$45. The idea here is that it would be unreasonable for either of the persons involved to simply retain their original opinion (Kelly 2009, 112). But it would also seem to be just as unreasonable for either of the persons to completely defer to the other person's opinion and abandon their own. It has been suggested that what is actually required in symmetrical cases where people are equally competent and equally reasonable and yet they disagree about a shared body of evidence, is that the opinions of each person should be given equal weight in each person's consideration of how to respond.

In the literature there are a number of approaches which at root believe in the symmetrical treatment of the disagreeing peers. They can conveniently be referred to as *conciliatory approaches* (Elga 2007a; Matheson 2009). According to these approaches if it is a given that peers have an equal probability of being correct, this is taken to entail that unless there is reason to think otherwise, their views ought to be given equal weight in assessing what is true. Adam Elga (Elga 2007a), David Christensen (Christensen 2007), and Richard Feldman (Feldman 2007; Feldman 2006), have all, to a certain degree, proposed symmetrical treatment of the beliefs of peers following the discovery of disagreement. Each of the views is somewhat different, but at root they all propose treating the beliefs of peers following disagreement symmetrically. For instance, according to Adam Elga's Equal Weight View (Elga 2007a), upon the discovery of disagreement you should give equal weight to your belief and to that of your peer. In what follows we shall refer to the views that treat the beliefs of disagreeing peers symmetrically, as conciliatory views, and we shall, at times, treat Elga's Equal Weight

View as a representative of the conciliatory approach. Elga's precise formulation of the Equal Weight View is as follows:

Upon finding out that an advisor disagrees, your probability that you are right should equal your prior conditional probability that you would be right. Prior to what? Prior to your thinking through the disputed issue, and finding out what the advisor thinks of it. Conditional on what? On whatever you have learned about the circumstances of the disagreement. (Elga 2007a, 490).

Now if in the restaurant example both opinions are given equal weight in deliberation, as the conciliatory approach suggests, whether \$43 is *your* result should have no bearing on the weight you assign to it in revising your belief. Additionally, the risk that you will grant your own convictions a privileged epistemic status is mitigated by the base-rate considerations of the equal probability of both answers being right. Hence because you are aware of your equal likelihood of being right, after finding out that your friend believes \$45, you ought to give the same weight to his opinion and to your own opinion in considering your response.

This suggests a symmetrical prescription regarding peer disagreements, and if one thinks of belief in all-or-nothing terms then in the case described above it may actually lead to a suspension of judgment about what the share of the bill actually is. There appear to be three assumptions that govern this symmetrical treatment of both beliefs. These are: 1) that each person holds a different belief about the share of the bill, but 2) each also knows that there is only one correct answer, and 3) each also thinks that prior to the calculation

either person was equally likely to get it. Given these three conditions, and in the absence of additional information about the circumstances, a symmetrical treatment of both beliefs is deemed to be the appropriate response.

Here is another related example, this time a depiction of a disagreement about matters philosophical, taken from Peter van Inwagen (van Inwagen 1996). This example is of a different sort to the restaurant bill example but it also characterizes incompatibility of beliefs in relation to the same considerations:

How can it be that equally intelligent and well-trained philosophers can disagree about freedom of the will or nominalism or the covering-law model of scientific explanation when each is aware of all of the arguments and distinctions and other relevant considerations that the others are aware of? ... How can we philosophers possibly regard ourselves as justified in believing much of anything of philosophical significance in this embarrassing circumstance? How can I believe (as I do) that free will is incompatible with determinism or that unrealized possibilities are not physical objects or that human beings are not four-dimensional things extended in time as well as in space, when David Lewis - a philosopher of truly formidable intelligence and insight and ability - rejects these things I believe and is already aware of and understands perfectly every argument that I could produce in their defense? (van Inwagen 1996, 138).

van Inwagen suggests the beginning of an answer to his own question, and it seems a fitting description of a common style of answer to this type of question:

Well, I do believe these things. And I believe that I am justified in believing them. And I am confident that I am right. But how can I take these positions? I don't know. That is itself a philosophical question, and I have no firm opinion about its correct answer. I suppose my best guess is that I enjoy some sort of *philosophical insight* (I mean in relation to these three particular theses) that, for all his merits, is somehow denied to Lewis. And this would have to be *an insight that is incommunicable* - at least I don't know how to communicate it, for I have done all I can to communicate it to Lewis, and he has understood perfectly everything I have said, and he has not come to share my conclusions (van Inwagen 1996, 138).

So, van Inwagen recognizes that Lewis is just as likely, if not more so, to see the truth in these matters. Yet van Inwagen and Lewis disagree, and van Inwagen is still assured his is the correct view. If this assurance is not to be groundless, he must attribute some sort of special philosophical insight to himself, and because his failure to persuade Lewis of his views persists, this insight looks to be incommunicable.

Something is descriptively correct in this account, if only because van Inwagen is pointing out that he has to believe in his own beliefs. However, recognizing that Lewis is his epistemic peer, he grounds this belief of his in his own beliefs in an “incommunicable philosophical insight” which is just the sort of consideration that would seem to be unlawful to use one-sidedly when considering how to respond to peer disagreement. If in the context of the present example we assume that Lewis is considered a peer (and not, as van Inwagen seems to modestly suggest, a philosophical superior), then “philosophical

insight” would be unlawful to use one-sidedly in a case where the philosophical considerations are shared, because it in essence supports a conclusion from which it derives. The justificatory force of philosophical insight in this context relates to how compelling the truth of a proposition is and so it cannot also be taken to support the truth of the proposition from which it derives its justificatory power. What is distinctive about van Inwagen’s example is that it suggests that even when philosophers are aware of and understand each other’s considerations, they can still come to hold different positions.

Consider that if two persons are equally likely to be right about philosophical matters (let us call such persons *philosophical peers*) then supposedly the conviction that each of them has about their belief should be given no weight at all in considerations of how to revise beliefs following disagreement. According to the Equal Weight View, if people have an equal probability of being right they should give the same weight to their conclusions, no matter what these conclusions happen to be. And arguably, this stipulation may also be taken to imply that the conviction that each individual may have in support of one of the conclusions being true has no justificatory weight. Hence in both the restaurant bill example and the – rather different – philosophical disagreement example, symmetry in competence leads to an equal weighing of beliefs which consequently makes any appeal to personal conviction irrelevant.

Let us now briefly go over two more examples adapted from the kind of cases presented in the literature. My aim in presenting these versions of peer disagreement is to begin arguing that they are importantly different from one another in such a way that suggests that a uniform response to all such cases of disagreement may not be plausible.

2.5.1 DRUGS AND ABORTIONS – TWO MORE EXAMPLES

I think you are highly intelligent, well informed, and morally sensitive. We are out walking and we discuss whether Class A drugs such as MDMA ought to be legalized. I – surprisingly – discover that you think that they should. Now let us ask: ought the fact that you believe that MDMA should be legalized to lead me to modify my belief that it shouldn't be legalized?

Well, yes, it should: *if* I expect from you what I expect from myself and you then respond differently to me, I should seemingly grant your – different – response a level of plausibility equal to mine, as the Equal Weight View suggests. But something seems wrong here. Since to grant your response a level of plausibility equal to mine, I would be granting plausibility to a belief I regard as clearly wrong. But in this case I regard it as wrong not because I feel sure about how I arrived at my belief, but because legalizing MDMA goes against some basic values and ideals that I have, in general.

The epistemic status of these basic values is somewhat analogous to van Inwagen's philosophical conviction. But although they are analogous to that conviction, basic values and a sense of philosophical conviction are importantly different. Basic values as I am referring to them here are interlaced with many other beliefs that I possess which I regard as expressive of these values. In this sense, embracing a belief that goes against a conviction about a particular philosophical problem is different to endorsing a view that is counter to one's basic values. In accepting a belief that goes against what one believes in other areas appears to entail that one must give up much more than with other beliefs that are not as heavily intertwined with one's other doxastic commitments. At first glance

the appeal to personal factors such as the personal cost of accepting a view that is different to other views one has does not seem permissible, because it is a personal concern and not an epistemic concern. Yet it is important to see that there is also an epistemic issue here; specifically, the epistemic issue relates to the epistemic cost of accepting a belief that does not accommodate to other beliefs which one is committed to.

Now, imagine that we both also happen to be of similar moral and religious disposition and of a similar educational, social, and economic background. Imagine that we also both know this about each other and that we are both also aware that in the past we have voiced the same kinds of opinions on religious and moral matters. In fact we take each other to be somewhat like moral peers, and justifiably so. On a particular occasion I hear that you believe abortions are morally impermissible, while I am firm in my belief that they are permissible²⁴; or, if it seems too unrealistic for two people of such similar disposition to hold opposing views, I hear that you believe that abortions are always morally impermissible, while I believe that they are sometimes permissible. In this situation, how should I respond upon learning what you think?

Just as in the previous examples, in this case too, we regard each other as equally capable in judging correctly in the moral domain²⁵, an area that past acquaintance with each other's responses has led us to count each other as peers. Now for me to suspend judgment on what I think about the permissibility of abortions or alternatively to recede my confidence in my own belief to an extent equal to the prior probability that I would

²⁴ The abortions example is adapted from Elga (Elga 2007).

²⁵ It is questionable whether the disagreement about MDMA belongs in the moral or legal domains. I think it could belong to both, depending on the type of disagreement surrounding MDMA.

have been correct would mean that I would be standing indifferently towards a belief I hold dearly, namely, that abortions *are* permissible, and I would be endorsing, at least to an extent, a belief which I firmly *do not* hold, namely, that abortions are impermissible.

The drug permissibility example and the abortions example are supposed to represent different types of cases in which peers disagree. Because both examples share what seems to be a common problem – i.e., two peers that disagree about the same matter – both cases, on the uniform approach to peer disagreement, are thought to also share a common solution for an appropriate response – for instance the Equal Weight View. But as implied above, it doesn't seem so obvious that a conciliatory response will be equally plausible in these different instances of peer disagreement. My contention is that while being an epistemic peer might *mean* the same thing in these different domains, it can *entail* different responses in different domains. This is the subject of the next section.

3. WHAT'S THE PROBLEM WITH A UNIFORM SOLUTION TO PEER DISAGREEMENT? (AND THE PRESENTATION OF MY ARGUMENT).

In this section I will argue that different responses to disagreement are appropriate in different contexts. What I will suggest is that typically, a disagreement about proposition p belonging to some domain Σ , which occurs between A and B, who are both epistemic peers in Σ , may make a particular response to disagreement – for instance the Equal Weight View - more plausible than it would be if A and B were to disagree on proposition q in the context of domain λ , a domain in which they also happen to be epistemic peers but which solicits a different response.

While the Equal Weight View is a frequently discussed response to peer disagreement, it is, as noted, one of a number of conciliatory responses thought to be appropriate following the discovery of disagreement.²⁶ What these conciliatory views have in common is that they contend, to varying degrees, that discovering disagreement requires that one make doxastic conciliation (Matheson 2009), which means that after discovering that a peer disagrees one ought to reduce confidence in one's belief. Some have suggested that this position, more specifically, the Equal Weight View, is unsatisfactory, and go on to suggest alternative approaches to responding to disagreement.²⁷ Nonetheless, the approaches that have been advanced in the literature, including those that propose alternatives to the Equal Weight View, seem to approach peer disagreement as a uniform phenomenon that requires a single revisionary response that is intended to resolve every instance of peer disagreement²⁸. These accounts seem to overlook the possibility that different domains make revisionary responses to disagreement varyingly plausible. It is because of this that I think that their underlying approach is wrong. I think that what makes it wrong is that their approach doesn't take into account the different ways in which beliefs in different domains are formed and the types of processes involved in the resolution of disagreements. This point may presently seem unclear but it will become clear in the course of my argument in the following pages.

²⁶ In addition to Elga (Elga 2007), Christensen (Christensen 2007) and Feldman (Feldman 2006) advance conciliatory responses to disagreement.

²⁷ E.g., Kelly's Total Evidence View (Kelly 2005; Kelly 2009) and Enoch's various objections and his own Common Sense View (Enoch 2010).

²⁸ Kornblith (Kornblith 2010) is perhaps an exception. I will discuss Kornblith proposal in the last part of the present section.

3.1 EPISTEMIC DOMAINS

Epistemic domains, such as arithmetic, philosophy, chess, medicine, morality, and any other areas in which people hold beliefs, vary not only by subject matter – e.g., arithmetic versus philosophy versus chess – but also in terms of the belief-forming procedures that characterize them. By a domain's belief-forming procedures, I understand those procedures that establish what the evidential basis for establishing true beliefs is, together with what the rules for epistemic conduct are, within that domain. Let me stress that I am not assuming that it is always obvious which rules apply where or that it is always uncontroversial what the rules are. I am simply stating that rules for epistemic conduct do vary by domain and that these rules are what are brought into play in, for instance, justification of why one believes as one does. These belief-forming procedures seem to me to be the definitive markers for distinguishing one epistemic domain from another in matters relating to what is involved in attaining true beliefs. And it is the varying epistemic characteristics of different domains that will serve as the foundation for my argument.

In some domains, such as arithmetic, there are clearly defined, explicit and transparent consensual procedures for arriving at beliefs. We often refer to procedures of this kind as 'rules'. But this is not to say that all procedures for arriving at beliefs can be adequately described as rules, nor that all procedures are consensual. What characterizes these kinds of rules is that they provide a relatively clear and straight-forward procedure for importing conclusions from the evidence. It is because of such rules that we know which procedures to follow and how they should be followed, when deliberating in domains such as arithmetic. A related issue here concerns the genesis and legitimacy of how

different belief-forming procedures and rules of epistemic conduct come to characterize different domains. This is something that I will discuss in section 3.3 in comparing conciliatory epistemic conduct in the judicial domain and in the domain of chess, domains in which belief-forming procedures differ significantly.

In other domains, such as for instance the domains of morality and religion, one might be confident that one has arrived at a moral or religious belief in the appropriate way. That is, according to the appropriate belief-forming procedures in these domains. But there won't always be consensus about what the appropriate procedures for forming moral or religious beliefs are. Nor will these procedures always be clearly articulated and transparent, which may make it problematic for different individuals to follow them in the same way.

To see this difference note that the domain of algebra has explicit and clear procedures for resolving disagreements about algebraic equations – namely, algebraic rules. Yet in the domain of philosophy, to take another domain in which the belief-forming procedures are different for different individuals, the way in which one would resolve divergence over a philosophical matter such as whether there are any synthetic *a priori* truths, or whether free will is in fact consistent with determinism, would be quite different for different people, and may also be different for the same person across different questions. This is not to say that philosophical truths are relative, but rather that the means of arriving at philosophical beliefs is not as clearly defined or universally agreed upon as arriving at beliefs is in other domains. To clarify: one can certainly identify a good philosophical argument and a sufficient philosophical proof; that is, the standards (or rules) for what is a good proof are fairly well agreed upon. But this does not mean that

there always is an easy to follow procedure for devising philosophical arguments or for arriving at answers to philosophical questions.

Additionally, the processes by which disagreements in domains other than philosophy – such as legal or moral domains - are ordinarily resolved are far less precisely defined and clearly determined than the way disagreements over algebraic equations are resolved. Consider for instance how disagreement about the innocence of an accused person might be settled between jurors in court. Or the way in which discussions about whether abortions are permissible would be settled in argument (Elga 2007a).

Let us look more closely at these two cases. The concepts that feature in those processes supposed to guide reasoning in court on matters of innocence assume a juror's ability to identify when there is 'doubt' or when there is 'inconclusive evidence.' While these concepts may be clearly defined and comprehensible to the juror, the procedures for implementing them are less precisely articulated. The epistemic complexity of an evidential situation or the limited availability of evidence may call for the juror to make calls on the basis of indeterminate and insufficient foundations. In a similar way, debating whether abortions are permissible would typically involve reasoning by way of subsidiary arguments that support or rebut particular views about various other claims such as the status of the fetus as a person. And the degree of support and the strength of rebuttal would in each case seem to be convincing for different people in different ways, depending on the subsidiary arguments they use and depending on the views with which they buttress their reasoning. Unlike the above mentioned clearly determined and articulated epistemic domains, in these kinds of legal or moral cases it is far from clear

what the belief-forming procedures mean and consequently how they should be implemented.

Some of the points I have raised in relation to legal and moral cases can also be applied to other areas where procedures are not always consensual or transparent, such as for instance certain areas in empirical science. While validation methodologies and belief-forming procedures may be clearly defined, explicit and agreed upon in experimental surroundings, the effective implementation of these procedures in the process of establishing scientific statements normally requires more than is articulated by the methodologies themselves.

3.1.1 *HOW COARSELY GRAINED ARE THESE DOMAINS?*

Before proceeding I would like to briefly attend to what appears to be a pressing concern. The way in which I have set up the claim that there are epistemically relevant differences between domains may have painted a somewhat simplistic picture. This, I think, is partly in view of the sorts of examples that I have chosen to represent the respective domains between which differences are apparent.

I think that it is uncontested that there are clearly and immediately identifiable differences between assessing truths in algebra and doing so on moral or in legal issues. But the question I would like to presently consider is how finely or coarsely grained is the differentiation that I have proposed, and what is its actual epistemic significance? Is, for instance, all mathematical knowledge to be taken together, and all moral knowledge to be taken together? Are basic arithmetic and topology part of the same epistemic domain? Are meta-ethics, general normative ethics, and particular moral judgments all part of the

same epistemic domain? Perhaps more importantly, could the encompassing high-level domain names that have been referred to lead us to neglect the genuine possibility for reasonable peer disagreement in mathematics by taking simple arithmetic and algebra examples as representative of the mathematical domain as a whole? Consider other mathematical areas in which reasonable peer disagreements may seem more plausible, Golbach's conjecture for instance, or preferences for Euclidian or non-Euclidian geometry²⁹.

These seem to me to be significant questions. Constructively, I believe that some domains, such as mathematics, possess their own spectrum of sub-domains, some of which appear to accommodate reasonable peer disagreement better than others. And as such it may be difficult to immediately accept encompassing revisionary solutions to different 'domains.' All the same, it doesn't seem immediately apparent that all domains contain the same degree of epistemic variability between their sub-domains. It isn't, for instance, immediately clear that the ethical or religious domains hold the same variety of epistemically distinguishable sub-domains as mathematics. What I propose is a way of meaningfully distinguishing between different epistemic domains by regarding them according to the manner by which they lend themselves to consensual epistemic conduct, deviation from which can more easily be identified.

3.2 A SCALE BETWEEN TWO EXTREMES

What I am proposing is that we take the belief-forming procedures of different epistemic domains as the axis on which to consider the appropriate responses to disagreement

²⁹ I am grateful to an anonymous referee for this point.

within those domains. As I have suggested, different domains have different forms of epistemic conduct – different rules, procedures, and methodologies for acquiring proper beliefs. In light of this I would like to suggest that epistemic domains should be thought of as located on an epistemic scale between two poles. Let us envision that at one end of the scale there are domains in which belief-forming procedures are clearly articulated, unambiguous, commonly recognized, and easy to follow. And on the other end of the scale there are domains whose belief-forming procedures are not as clearly articulated or commonly recognized, and that what the evidence should be taken to support in these domains is less clear.

What I want to suggest is that whether or not a particular response to peer disagreement is plausible will depend on where on the epistemic scale the particular disagreement is located. Roughly speaking, my contention is that the closer the domain in which the disagreement occurs is to the clearly articulated and unambiguous end of the scale, the more conciliatory a response to disagreement ought to be, perhaps even to the point of accepting egalitarian approaches to peer disagreements and splitting the difference between our opinion and that of our peer.

Let me try and convince the reader why this is so. Firstly, consider that each of us seems to have different intuitions about what the appropriate response to disagreement is in different domains. And in those domains in which the procedures for epistemic conduct are not clearly defined, or where there isn't consensus about what the relevant procedures for arriving at true beliefs are there seems to be a greater scope for an individual to judge how to respond to disagreement. By contrast, in cases where the procedures for epistemic conduct are clear and largely consensual there would appear to be less room for an

individual to judge what ought to be the appropriate response. Secondly, it seems that the more a belief relates to other beliefs the greater the cost that giving it up seems to incur. Moral or legal cases seem to be a good example here. If, following disagreement, I would need to back down from my belief that MDMA should not be legalized, this would seem to require me to recede from a variety of other beliefs that I have which appear to support this position. These include the belief that people should not be allowed to choose whether to use a hallucinogenic drug; that allowing people to choose whether to use the drug is an affirmation that the use of the drug is OK; that conveying that the drug is OK will lead people to use it more often than they use it when it is not permissible; that more people using MDMA will have an all-round negative effect. It seems to be the case that the less clear and unambiguous the belief-forming procedures of an epistemic domain, the more one seems to have to rely on the other beliefs that one has. Thus, on the occasion of disagreement in such a domain, there is a greater cost incurred by giving up beliefs for the sake of doxastic conciliation.

I do not have a firm diagnosis of what makes it true that conciliatory responses are more plausible as responses to disagreement at the clear and unambiguous end of the epistemic scale, but I would like to suggest that it relates to the role of individual judgment within the belief-forming procedure in these domains. To see this, let us make a comparison of a case from both ends of the scale. First we shall attend to a Jury case and then, in a moment, to a case relating to the game of chess. By contrasting these two cases I shall make the point about the role of individual judgment.

3.3 WHY CONCILIATORY RESPONSES ARE MORE PLAUSIBLE AT THE CLEAR AND UNAMBIGUOUS END OF THE EPISTEMIC SCALE

Consider the role that individual judgment has in deciding whether there is reasonable doubt about an accused person's guilt. While it may seem to be a recognizable and consensual guideline for judicial behavior, reasonable doubt relies heavily on an individual's ability to identify what *counts* as reasonable doubt and whether there *is* reasonable doubt in a particular case. Conversely, consider the role of individual judgment in deciding whether a particular chessboard is in fact three moves from checkmate. What is involved in this latter case undoubtedly requires less in terms of individual judgment regarding the nature of the appropriate application of the belief-forming procedures in chess. And it certainly leaves less room for idiosyncratic understanding of what things such as 'three moves' are, or what 'checkmate' is, both of which are clearly defined by the rules of chess. It is because of this that I believe that the rules of chess, just like the epistemic procedures of domains of a comparable unambiguous nature, leave less space for an individual to judge how to implement the appropriate belief forming procedures, than would be the case when doubting someone's guilt. Furthermore, because a domain such as chess leaves less margin for judgmental intervention in the application of its rules, one's justified confidence in having arrived at a conclusion correctly – that is, having followed the unambiguous rules of chess – would seem to allow less room for whim, intuition or individual interpretation to take over, and this seemingly leads to greater confidence in having calculated the moves of the game correctly.

We might understand the situation as one where different domains have different justified confidence levels that contribute to the level of plausibility of varying responses to peer disagreement. Typically, the beliefs that are held within a particular domain provide

justified confidence to lesser or greater degrees, depending on the available evidence and on the intrinsic epistemic characteristics of each domain, such as for instance their belief-forming procedures, which determine the type of inference made from evidence to beliefs. And what I am suggesting is that the more clear and unambiguous the epistemic procedures of a domain are, the more confident you will be that in acquiring your belief you followed the right rules and did so correctly.

Note that I am not claiming that in domains where the epistemic procedure is clear and unambiguous you will be more confident about your belief *per se*, but rather that you will be more confident that you have arrived at your belief *correctly*. That is, you will be more confident that you arrived at your belief by the appropriate belief-forming procedures that befit the domain and support the possession of true beliefs in that domain. To stress - this is not to say that you will in general be more confident about mathematical beliefs than about judicial beliefs, but rather that you will typically be more confident that you have arrived at your mathematical belief by the procedure that is appropriate for the mathematical domain. This does not of course exclude cases in which you use the right procedure but misapply it, which in most cases seems to be a genuine risk. In this sense I am claiming that in most epistemic domains and most knowledge forming methods within them, there is probably a different level of risk, but always some risk, across all the possibilities – e.g., less risk doing arithmetic than doing complex calculus.

To illustrate this consider the difference between arriving at an integer after having considered a simple algebraic equation, and arriving at an opinion about a preferred candidate for presidency after having considered the contenders. I take this to reflect the difference between well-regulated domains to less-well regulated domains. While you

may be equally convinced of your resulting belief in both cases, it should be clear to you that the epistemic foundations on which your opinion is based in relation to the presidential candidate are far less articulated, or at least far less explicit and transparent, than those in the algebraic domain. And while you might be equally confident in your belief in both cases, in the algebraic domain you will ordinarily be more confident of having applied the right procedures in establishing your belief.

At first glance, after discovering conflict in an area in which you feel confident about having formed your beliefs correctly, you should be less prone to conciliation as a way of resolving issues of disagreement. In such cases there is always a clear something you can point to, to support yourself. If, for instance, you have arrived at an algebraic conclusion you may, if you are sufficiently competent, know that you followed explicit and clearly articulated rules of inference and that generally speaking being wrong about such mathematical results comes down to performance errors in applying the rules, not to differences in judgmental appraisal. However, because you feel justifiably confident in your result in the mathematical case, your peer is seemingly just as justifiably confident in his result as you are because he is assumed to be equally competent in applying the relevant rules. And as long as the belief-forming procedures are clear, explicit, consensual, and easy to follow, disagreement between the two of you is taken to be the result of performance errors in applying the rules. And it is the equal likelihood of encountering performance errors that seems to also be implicitly entailed by the qualification of epistemic peerhood; this at least is what you ought to believe. And therefore it is because the rules are unambiguous that a strict approach to revision following disagreement is all the more forthcoming in accommodating your equal prior

probability for going wrong. And this initial confidence in the belief-forming method is what seems to make conciliatory views such as the Equal Weight View seem more plausible in domains where the primary reason for disagreement comes down to performance errors in applying the rules of inference within that domain.

For the same reason, domains in which epistemic conduct is not governed by clear and explicit belief-forming procedures, will not naturally appeal to performance errors in applying the rules as reasons for divergence. In matters pertaining to areas such as morality, politics, and various kinds of social judgments, it is judgmental appraisal that does most of the work in forming beliefs, and it is judgmental appraisal and the reasons it provides which typically support beliefs in these domains. Hence the natural explanation to appeal to upon encountering conflict with a peer in such cases is that the difference in belief has been caused by a difference in judgmental appraisal. And this kind of difference makes conciliatory approaches less plausible. When each of the peers explains the difference in belief as a difference in judgmental appraisal this gives all the more reason to count the disagreement as nullifying the relevance and consequently the applicability of the previously established symmetry. To put this in clear form: if a disagreement is explained by performance errors in applying rules of epistemic conduct then the conciliatory approach seems more appropriate; if a disagreement is explained by differences in judgmental appraisal, this reason in itself reflects an essential rather than a contingent difference. This point is worth stressing since it appears to go against a view advanced by Hilary Kornblith (Kornblith 2010), according to which we should reserve judgment in exactly those areas – such as philosophy and morality – that are marked by

persistent divergence in overall outlook. I will briefly attend to Kornblith's position and emphasize how it differs to the view I am advancing here.

3.4 INTERLUDE: KORNBLITH AND THE CONSENSUAL FOUNDATIONS OF CONCILIATION

Kornblith (Kornblith 2010) also proposes looking at the viability of conciliatory responses to peer disagreements in relation to the domains in which these disagreements occur. But for Kornblith it is not the nature of epistemic conduct within a domain which determines the plausibility of conciliation, but rather whether or not beliefs within the domain have a successful track record of progressing toward consensus. Kornblith focuses on the question of whether philosophical disagreements between philosophical peers merit conciliation. But his question generalizes and applies to other domains in which the beliefs of experts, just like those of philosophers, do not always converge to consensus.

The heart of the claim is that in domains such as certain domains in the empirical sciences, where the opinions of experts tend to progress toward consensus, the beliefs of experts are epistemically justified and can be relied on. This is because historically, there is not too large a divergence between the beliefs of experts in these domains, and one can assume that what is right to believe is what most experts believe. And therefore when there is a case of experts disagreeing with one another, both are epistemically justified and reliable, and hence conciliation between their views is appropriate.

Yet in philosophy there doesn't appear to be a consensual track record between philosophers, or any kind of advancement toward one, and "without that background of longstanding progress, we must look at individual investigators quite differently"

(Kornblith, in: Feldman and Warfield 2010, 51). Kornblith suggests that because of this lack of track record of consensual progress, philosophical opinions - whatever they are – aren't epistemically justified, because there is no consensual standard for establishing the reliability of experts to go by. And this generalizes to moral and political issues too³⁰. Kornblith's concludes that in philosophy and other domains where there is no significant track record of consensual expert opinion, "we must, in the end, withhold opinion on the issues under consideration," and suspend our judgment on what we are disagreeing about.

Above I claimed that differences in judgmental appraisal make conciliatory approaches less plausible. I also noted that judgmental appraisal plays a greater role in some domains than in others. I noted that in those domains where judgmental appraisal is less intrinsic to the belief-forming process, such as in mathematics or in chess, disagreements come down to performance errors, which support conciliatory solutions. And where judgmental appraisal is more intrinsic to the belief-forming process, such as in philosophy and morality, it is less natural to explain disagreements as resulting from performance errors. In fact, in such cases, it is not strictly speaking the application of a rule that supports a belief, but rather a more entrenched and elusive process which one is less likely to give up, both for the sake of conciliation and for the suspension of judgment. I will say more about this in the following section.

3.5 BACK TO CONCILIATORY RESPONSES

³⁰ "In philosophy ... there is no ... history of longstanding progress, and for that reason we should not consider the experts in the field – including, of course, ourselves, to be highly reliable" (Kornblith, in: Feldman and Warfield 2010, 52).

I shall now argue why conciliation is an appropriate response in the domains I have suggested. In the upcoming discussion I will suggest the different sorts of domains there might be according to the characteristics that place a domain at one point on the scale or another. But I will also discuss the different responses to disagreement that are appropriate in each sort of domain.

Explaining a disagreement between peers as due to performance errors seems to make a conciliatory response more plausible. As long as peers have the same probability of performing well in a particular area, they also have the same probability of performing badly in that area. And performing badly can be taken to mean different things in different domains. In domains in which the procedures for epistemic conduct are clear, articulated and consensual, performing badly will largely be explained by performance errors – that is, by a failure to apply the right rules where they are relevant. In domains in which procedures for epistemic conduct are more ambiguous and less consensual, performing badly may mean a variety of other things that may have caused someone to go wrong. Because individual judgment plays a more dominant role in these domains, disagreements will often be explained by errors of judgment or judgmental appraisal rather than performance errors in applying the rules correctly. In demarcating the different sorts of domains across the epistemic scale I have suggested that some domains are easy to describe while others may not be so easy to place on a scale or alternatively, their essential epistemic properties are not easy to describe. Chess, for instance, is very much rule-governed by explicit rules, and so can be categorized precisely on scale. But other domains are not conventionally rule-governed and thus cannot be described and demarcated as easily.

This, then, is my argument to this point. Different epistemic domains differ not only in the nature of the objects that concern them, but in the types of rules that are available for forming epistemic judgments. Differences include the objective status of the rules, the role of personal judgment or skill, the level of background assumption invoked, and so forth. One crucial characteristic of a domain is how clearly articulated and generally accepted its belief-forming procedures are. Domains with very explicit rules lend believers greater confidence in beliefs founded on having followed the appropriate rules. In such domains, when two peers disagree, conciliation is a very plausible response. This is because the performance of both peers seems somehow objectively quantifiable - a difference is evident, no difference can be determined, and so the doubt must distribute evenly across both believers. There are other types of epistemic domains, however, where the rules are less clear or explicit, perhaps more subjective, and so this argument in favor of conciliation is not as obvious. I will presently look in more detail at some of these cases. But before I turn to do so I would like to propose an additional metrics for evaluating the epistemic characteristics of domains. Aside from whether or not their epistemic procedures are clearly articulated, which may be regarded by some as perhaps, rather loose, it might be possible to regard domains according to their susceptibility to performance errors. In this sense the type of response appropriate upon discovery of disagreement will depend on how susceptible the domain in which the disagreement occurs is to performance errors in applying the sort of epistemic conduct that such a domain requires.

3.6 WHEN CONCILIATORY RESPONSES ARE LESS NATURAL

Take a case where you and I disagree about whether or not *this* chess position is three moves from checkmate. If I know that I know the rules of chess, and I know that you know the rules of chess, and I know that the ability that is needed for either of us to know that this chessboard-state is three moves from checkmate is the same ability in each of us, and, let us say, involves being able to think forward and apply the rules of chess three moves ahead, then I also know that the likelihood that you strayed from the rules in arriving at your belief is small, just as small as the likelihood that I have strayed from the rules. Why is this important? It is important because if you did stray from the rules of chess, such as, for instance, by allowing the Rook to move like a Knight, then you did so because of performance errors, which essentially means that you implemented the rules incorrectly. And it is because, as peers, you and I are equally likely to get things right, we are also equally likely to get things wrong. And in domains where the belief forming processes are explicit and clear, going wrong amounts to going wrong in applying the rules. And because I must assume that we are equally likely to have implemented the rules incorrectly, this makes my conciliatory approach to disagreement all the more plausible.

But because forming beliefs in domains that are more epistemically complex requires more by way of individual judgment in the formation of beliefs, it is less natural to attribute difference in belief between disagreeing parties to errors in implementing the rules, simply because the rules have a less definitive role in the formation of beliefs in these domains. And by implication this is different to epistemically straightforward domains which require less judgmental intervention. And it may be more natural to

attribute differences in beliefs in such cases to differences in individual judgment between the disagreeing parties. And as long as disagreements are explained by differences in individual judgment, this makes base-rate considerations about forming beliefs correctly less relevant.³¹ Once base-rate considerations about equal prior probability become perceived as less relevant, the conciliatory responses that seem to be implied by such considerations become less plausible.

Moreover, in domains where individual judgment has a greater role in forming beliefs, it also has a greater role in justifying these beliefs. And when one is required to abandon a belief that was largely shaped by one's individual judgment, this may have a greater individual cost, since it would amount to disregarding individually defining factors that contribute to judgment – such as values, instincts, and intuitions. And the cost of disregarding these kinds of factors may not only be substantive – that is, valued in relation to the abandonment of the relevant values, instincts, and intuitions – but also symbolic – that is, valued in terms of the meaning of the abandonment of such individually defining factors.

If, for instance, we do not agree about whether a particular political approach is just, or whether a philosophical argument is convincing, the reasons that each of us appeals to when considering whether to revise our beliefs following our mutual encounter with each other's conflicting opinions are far more reflective of our personal persuasion than they are of any domain-specific rules of conduct that can be articulated. And this, I contend,

³¹ For empirical research suggesting when it is that base-rate considerations are taken into account in reasoning see: (Bar-Hillel 1980; Bar-Hillel and Fischhoff 1981; Gigerenzer, Hell, and Blank 1988; Payne and University 1976, chap. 8; Kahneman and Tversky 1973).

makes forsaking our own beliefs in the spirit of conciliation all the more implausible. Earlier we alluded to the interface of a disputed belief with other beliefs as having a similar ripple effect. But what I am describing here is slightly different because it turns on the kind of reasons we appeal to instead of attending to which beliefs are compromised by adopting a particular view.

Before we move on to consider a possible objection to the view I am proposing, it is important to contrast conciliation as a response to disagreement at one side of the scale with the type of responses that are plausible at the other end of the scale, where domains are less regulated by clearly articulated rules of inference and generally more ambiguous in nature.

On moral, political and social questions, the rules of proper epistemic inference may be less explicit but this does not mean that individuals are less certain about what they believe in these areas. In fact, it is perhaps because the procedures for arriving at moral, political and social beliefs are not transparent and the convention for epistemic conduct in these areas not typically rule governed, that people may actually be less willing to back down from their beliefs. Because beliefs in domains which are not governed by explicit epistemic procedures depend on individual judgment, intuition, and non-explicit reasoning tactics, it is less plausible to talk of performance errors and misapplying the appropriate procedures, simply because there are none. It is more plausible to contend that conflicts in belief are evidence of disparity of individual persuasion rather than evidence of imperfect competence in applying conventionally recognized epistemic procedures. To put this more clearly: social and moral judgment appears to involve

individual judgment to a greater degree. But additionally, individual persuasion is also thought to be constitutive of many moral, social, and political beliefs.

4. PRACTICAL VERSUS THEORETICAL SOLUTIONS

Here is a possible objection to my argument:

The considerations that you raise about there being domain-specific belief forming procedures do not affect the plausibility of a particular revisionary response to peer disagreement. They might relate to the practical problem of peer disagreement but they do not have anything to do with the epistemic problem of peer disagreement.

Here is the concern expressed in this objection. The practical problem of peer disagreement presents a scenario where you can recheck your belief. The epistemic problem of peer disagreement which is the problem addressed in the philosophical literature, considers the response that is required in the interim between discovering the disagreement and being able to reassess the evidence. The philosophical question regarding disagreement may even be thought of as a question that is concerned with the response that is required when there is no possibility to revisit the evidence, or when revisiting the evidence leaves both sides with the same beliefs. The heart of the objection is as follows. The procedures that substantiate what counts as a true belief in a particular domain may indeed assist in resolving disagreements in that domain. Sure enough, in some domains – typically domains at the unambiguous end of the epistemic scale - resolving a dispute is something that can be done by simply verifying each of the beliefs

through the appropriate procedure – i.e., by implementing the appropriate rules of inference. But because the problem of peer disagreement that we wish to address in this debate does not actually relate to cases where the evidence *can* be reassessed, or rather because the question of whether evidence can be reassessed isn't relevant, the claim that the plausibility of a response varies by domain is in fact unfounded. Because, the argument continues, the argument for domain variability relates to the epistemic procedures that enable to re-check, re-think, and re-assess the evidence following the discovery of disagreement, it does not seem to be relevant for assessing the appropriate *epistemic* response to peer disagreement. Since whether or not a disagreement could be resolved by assessing the disputed beliefs by way of an epistemic rule or procedure is of no concern, since according to the way the problem has been presented in the philosophical literature, the contested beliefs cannot in fact be resolved, and this is why it is an epistemic problem and not a practical one. And hence the domain variability that has been expounded here is not relevant to the disagreement debate or to establishing the plausibility of particular responses to disagreement.

My response to this objection can be put concisely as follows. Firstly, while on first sight this might look like a reasonable criticism, it is based on a highly artificial hypothetical scenario, which will rarely, if at all, be instantiated. Secondly, the re-evaluation of evidence and the re-considering of what previously encountered evidence means are indeed practical possibilities. But additionally, they also serve as considerations in assessing the plausibility of epistemic responses to peer disagreement. Here are two reasons why I think this is so. Firstly, in addition to being practical concerns, considerations of how evidence ought to be re-evaluated in a domain also serve non-

practical purposes – that is, epistemic roles - which relate to the belief forming procedures in those domains. Typically, what is entailed by reflecting over the justificatory status of a particular belief involves consideration of what *would* go towards justifying that belief. That is, it involves a hypothetical consideration of the procedure by which such a belief ought to be formed, whether or not this procedure was the cause of the belief or its origin. And it is on the basis of this assumption that considerations for resolving doxastic disputes are also non-practical, epistemic considerations. The fact that different domains have different procedures for evaluating true beliefs becomes an epistemic consideration because if peers disagree in a domain in which the procedures are clear and unambiguous this becomes a reason against their continuing to believe as they believed prior to the disagreement. So the first point is that when disagreement occurs, doubt is cast on the very practices of belief forming, and this doubt causes one to evaluate the practices on which one's beliefs were formed. In this sense, it is right to say that there is no such thing as the pure epistemic problem which ignores these practical considerations. The pure epistemic problem would be very much a limit case where such methods are assumed.

Secondly, the fact that a person recognizes that a disagreement is about a subject that belongs to a domain that has explicit and mutually recognized rules of resolution will bear on the plausibility of a particular revisionary response because of how that person believes his and his peer's beliefs were formed, not necessarily because of his practical concern about how his own beliefs will be revised. Thus the second point is not about evaluating the practices but simply pointing out that to call someone a peer is to have said

you have evaluated his epistemic skill, so one cannot use the term peer without implying some practical consideration.

To conclude our discussion I suggest that disagreements encountered in different domains across the epistemic scale provide different kinds of reasons for sticking to our beliefs and changing our minds. I have indicated a beginning taxonomy of some of these different types of belief and reasons for holding them. And I have argued that these different kinds of reasons can make particular response to disagreement more or less plausible.

Abstract: this paper addresses a core topic in the debate about peer disagreement: whether, and if yes, how to revise one's views upon discovering that one is disagreeing with a peer – a colleague or fellow expert - who is regarded as being equally competent when it comes to the debated matter at hand. The topic is relevant in all wakes of life where shared information is at hand and different beliefs and opinions may be derived from it. My approach to this topic is non-standard. I argue that there is a difference between addressing this core question hypothetically or in theory as an abstraction of a case of disagreement, and this same question when it is considered from a practical point of view, in consideration of features that characterize *actual* cases of disagreement. The core difference, I suggest, relates to judgments that pertain to the relevance of second-order evidence (roughly, evidence about the viability of inferences made from evidence). In actual cases of disagreement, as opposed to hypothetical abstractions of such cases, what is or isn't relevant depends on how uncertain the situation is seen to be by the agent in question. And this depends on their level of confidence as perceived by subjective first-person judgment. I show that subjective confidence about first-order judgments can swamp second-order evidence – against a plausible view that it shouldn't. I regard this as a significant problem in practical rationality that is brought into focus by disagreement problems. I focus on the problem of relevancy judgments with regard to the calibration of

³² For a published version of this chapter, see: Martini, Carlo & Marcel Boumans (Editors), *Experts and Consensus in Economics and the Social Sciences*, Springer, *forthcoming*.

first-person judgments with statistical data. I show how this impinges on disagreements when these are considered from the first-person perspective.

1. INTRODUCTION

Recent debates dealing with the epistemic significance of peer disagreement have sought to provide responses to cases in which peers disagree about the epistemic import of a shared body of evidence. Various responses have been suggested in the literature³³. Typically, these address problems of the following general form:

If persons A and B are epistemic peers – meaning roughly that it is equally probable that A and B will be correct in the domain in which they are peers – and on a particular unexpected occasion they happen to find out that they disagree about whether a particular proposition P is true given the evidence that is equally available to them both, and assuming that neither party has any independent reason to discount the dissenting party’s conclusion, they ought to respond to this discovery ... *in such and such a way*.

The responses to this problem in the literature vary, and can, I believe, be divided into three kinds: 1) the *bootstrap* response; 2) the *conciliatory* response, and 3) the *egalitarian* response. I think that some of these responses have considerable appeal. But I also think

³³ Such as the Equal Weight View (Elga 2007), the Total Evidence View (Kelly 2009), the Common Sense View (Enoch 2010) as well as a number of other closely related approaches (Feldman 2006; Feldman 2007; Christensen 2007; Matheson 2009; Moss 2012).

that some make sense theoretically, but are not practically rational, and that others are simply unreasonable. All the same, I do not believe that the goal of establishing *which* of these responses is better, as a good deal of the recent debate has been focused on doing, actually addresses the normative problem of disagreement.

Customarily, the problem of disagreement asks about the appropriate response (typically the response of a peer) upon encountering a disagreement between peers. The literature offers different solutions to the problem, each of which has more or less normative appeal. Yet none of these solutions seems to engage with what seems to be the *real* problem of disagreement. It is my aim in this paper to highlight what the real problem of disagreement is. It is, roughly, the problem of deciding *whether* a revisionary tactic is appropriate following the discovery of disagreement, as well as deciding *which* revisionary tactic is appropriate. This non-standard approach to the disagreement problem exposes a slippery and inevitable difficulty that any discussion of disagreement ought to deal with. Once recognized, the real problem of disagreement reflects on the standard question addressed in the literature about *which* revisionary tactic is appropriate for dealing rationally with disagreement. The problem also generalizes to broader problems in practical rationality.

The above-mentioned approaches (1-3 above) are characterized by the different tactics that they propose for dealing with disagreement. But these tactics only appear to be relevant after the truly hard work of deciding *whether* they are relevant in each actual case of disagreement has been done. And this, I believe, is a huge problem that has not been adequately recognized in the literature or has even largely been missed until now. It

is, in a word, the problem of judgments about relevancy³⁴. The epistemic significance of this problem extends beyond debates surrounding disagreement. Relevancy judgments involve the subjective appreciation of evidence about the reliability of inferences from evidence. It is my aim here to draw attention to this problem which, I believe, also lies at the heart of debates surrounding disagreement. It is my contention that *actual* cases of disagreement, as opposed to *possible* cases of disagreement, must deal with this inevitable situation.

The paper will proceed in two stages. In the subsequent section I will outline what I take to be the real problem of disagreement, setting forth my core argument. But I will start with some preliminaries. After that I will present three approaches that I take to characterize the solutions that have been proposed in the literature. In the course of doing so I will show why most of these do not address the real problem of disagreement. But I will also suggest which of the approaches in the literature is most plausible in view of its partial recognition of the underlying difficulty brought about by the inescapability of relevancy judgments. My foremost aim in this paper is to highlight a fundamental difficulty relating to first-person judgments about the relevance of second-order information – roughly, information which is applicable to specific cases by virtue of their location in a broader statistical framework. I aim to demonstrate how this difficulty impinges on actual situations where revisionary responses to disagreement are called for.

2 THE REAL PROBLEM OF DISAGREEMENT – *SETTING THE STAGE*

³⁴ I have dealt with the problem of relevancy in reasoning in a separate paper.

The crux of the matter lies in a practical paradox of sorts, which, as I shall presently show, is inevitable. It relates to judgments about evidence made from subjective standpoints. Before I present this practical paradox I will start by clarifying what I mean by disagreement, to which, as I will demonstrate, this paradox applies.

2.1 DISAGREEMENTS BETWEEN ORDINARY PEOPLE

Disagreements as I shall refer to them are situations in which *people* disagree. More specifically, the types of disagreements I want to concentrate on, in the spirit of the recent literature, involve cases where one person finds out that another person, typically someone whose relevant epistemic capabilities are similar, holds a different opinion, view, or belief on the same matter³⁵. The fact that I will be referring to ordinary people in this context is important, and I make note of it here because it imparts that I wish to relate to the normative question of how *human* reasoners ought to respond to disagreements in situations that are characterized by uncertainty about who or what is correct. This approach contrasts to another, prevalent in the economic as well as the philosophical literature, in which rational agents rather than human reasoners are the focus³⁶. The rational behavior of rational agents is typically different to that of human reasoners, and so as to set the stage for addressing the problem, the distinction must be made.

2.2 A BRIEF NOTE ON DISAGREEMENTS IN THE ECONOMIC LITERATURE

³⁵ While opinions, views, and beliefs may suggest different meanings, each suited more than the other for a particular context; I use them here interchangeably as referring to what a person regards as true.

³⁶ By rational agents I have in mind something similar to what Thaler and Sunstein have recently referred to as *Econs*. See: (Thaler and Sunstein 2008).

In the economic literature there has been an ongoing debate since the mid-seventies surrounding the question of disagreement³⁷. The core of this debate focuses on the possibility of rational disagreement, or conflict, between rational agents. The question posed is whether it is possible that agents who are, broadly speaking, programmed to conditionalize on information in the same way, may *agree to disagree* (Aumann 1976). The debate considers whether it is possible for rational agents to disagree *rationally*.

The types of agents referred to in the economic literature dealing with questions concerning disagreement are *not* human agents. And the rationality that is attributed to these agents is perhaps *not* human rationality. Human rationality, in the context of disagreements, relates to human reasoners that encounter evidence to which they may respond imperfectly. The problem is located in the wider context of human fallibility and regards disagreements as opportunities for corrective measures aimed at mitigating erroneous consequences of imperfect reasoning.

2.3 HUMAN IMPERFECTION AND ITS IMPLICATIONS FOR PRACTICAL REASONING PROBLEMS

There are two critical senses in which human imperfection impinges on disagreements. The first relates to the gathering of information, the second to inferring conclusions from it. People's capacities in both these areas are limited, and they commonly make mistakes. In the course of my discussion I will assume that people are typically aware that they make mistakes; they know that they are not normally capable of taking all or perhaps

³⁷ See for instance: (Aumann 1976) and (John Geanakoplos and Heracles M. Polemarchakis 1982; Jonathan A.K. 1983; Moses and Nachum 1990; Rubinstein and Wolinsky 1990; Robin Hanson 2003; Dégremont and Roy 2009; Hansen and Cowen; Milgrom and Stokey 1982).

unlimited information into account when they are deliberating in uncertain conditions. Moreover, they are normally also aware of the fact that when they do possess information that is relevant to their beliefs, their responses - characteristically their inferences from this information - are often imperfect. And by 'imperfect' I have in mind, approximately, three things.

2.4 IMPERFECTION AND REASONING

Firstly, when people infer conclusions from evidence, however limited or encompassing this evidence may be, they are not always correct in what they infer. In addition, people commonly recognize this about themselves. And what this actually means is that they recognize that their reasoning is error prone, and thus imperfect.

Secondly, while people generally know that their reasoning is error prone, they do not always recognize the occasions in which it is so. Because of this, people often think they are right when they are wrong, and therefore incorrect reasoning sometimes goes unrecognized.

Thirdly, because they know that they may sometimes be wrong about what they believe, and because they also know that they don't always recognize the occasions when this is so, people should not always be certain that what they believe is correct. Indeed, it appears that in general they ought to have some reservations about the viability of their responses, particularly when they encounter dissent from an esteemed counterpart or a fellow expert.

2.5 WHEN SUBJECTIVE CREDENCE PLAYS AGAINST FACTS OF THE MATTER

In many situations of uncertainty, evidence may be more or less convincing, and this seems to play subjective credences against perhaps unknown facts of the matter. All the human reasoner has to go by is his subjective credence, which is assumed to mirror the strength of the evidence that he has. In cases such as this, an individual's subjective flaws as an evidence evaluator prescribe some type of risk mitigating strategy so that inferences whose impact on credence is partly subjective, can be weighed against some type of objective standard that is not based on the same error prone reasoning.

2.6 FIRST-ORDER AND SECOND-ORDER EVIDENCE

A helpful distinction has been made in the literature between two kinds of evidence³⁸. The distinction provides a convenient taxonomy for considering corrective standards for mitigating erroneous tendencies in practical reasoning. The distinction is between first-order evidence and second-order evidence. It can be understood as making a point about two kinds of epistemically relevant considerations, or evidence.

First-order evidence refers to the kind of evidence the presence of which can increase or decrease subjective credence in a particular proposition. And by increase, I mean that it can make one more confident about the proposition than before – either by moving one's belief towards that proposition, or strengthening it. In the same way, first-order evidence can decrease one's confidence too. Moreover, if first-order evidence is epistemically insignificant, it may neither increase nor decrease credence. The notion of second-order

³⁸ Kelly (Kelly 2009), Christensen (Christensen 2008), and Feldman (Feldman 2006) also refer to this distinction.

evidence refers to evidence that bears on whether one's beliefs, or the credences one has assigned, are likely to be correct³⁹. For our purposes what is entailed by first-order evidence is partly subjective because credences that are based on first-order evidence are estimates of a proposition's truth value, based on subjective assessments of first-order evidence. Credences might appear precise because they are usually expressed in numerical form, but let's remember that credences are people's estimates of truth values, expressed as probabilities. As opposed to first-order evidence which usually relates to a person's present judgment, second-order evidence is typically evidence that is based on past epistemic performance, or experience, and as such does not depend on corroboration by present judgment.

Here is an example illustrating this distinction. The first-order evidence (FOE) that I encounter may be the Candlestick in the Hall, which supports my belief that Colonel Mustard did it (P): $\langle \text{FOE} \mid \text{P} \rangle$, or else it increases credence in the belief that he did it. The second-order evidence (SOE) that I possess may be prior knowledge that in the past, when I inferred *who* was guilty on the basis of weapon and location alone (FOE), I was wrong 70% of the time. In this case second-order evidence is the knowledge that I have about my past performance in inferring conclusions in similar conditions (using the same variables)⁴⁰. It tells me how likely it is that my inference – e.g., $\langle \text{FOE} \mid \text{P} \rangle$, is correct.

³⁹ Compare Kelly's discussion of Downward Epistemic Push (Kelly 2009, sec. 5.3).

⁴⁰ Second-order evidence, conceived as I am presently suggesting, provides information about the likelihood of some event, outcome or possibility in some general population of events, outcomes, or possibilities. It may take many different forms and prior experience or performance is only one such form.

In the context of disagreements between peers, first-order evidence will be the evidence that each peer encounters and which consequently leads each to believe as he does in the first-place. Therefore if the disagreement is, for instance, between weather forecasters, and concerns the weather forecast for tomorrow, first-order evidence is the evidence on which each person bases his belief about tomorrow's weather. More generally, we might think of this as the type of evidence that is normally needed for a weather forecaster to make up his mind about tomorrow's weather, prior, that is, to finding out what his peer believes about it. Thus first-order evidence in this context may perhaps be temperature maps (TM), atmospheric factors (AF), and other metrological features (MF) on which weather forecasters typically base their predictions. Let E denote a particular piece of evidence. In this case a prediction based on FOE will look like this: $\langle \text{ETM}; \text{EAF}; \text{EMF} | \text{P} \rangle$.

Second-order evidence will typically be evidence that relates to the belief-forming circumstances in which conclusions are inferred from first-order evidence. This may for instance include considerations about how likely it is that the inferences made from first-order evidence are correct. In disagreement problems, second-order evidence will typically relate to prior knowledge about a person's competence in inferring conclusions from first-order evidence (typically stated in probabilities based on prior performance), or perhaps knowledge about the person's susceptibility to error.

The widely consensual position is that in reasoning problems that involve first and second-order evidence, subjective credences that are based on first-order evidence ought to be balanced by probabilities derived from second-order evidence based on past performance (in the same way circumstantial indicators in Bayesian reasoning problems

are weighed against base-rate information⁴¹). The general contention is that not to take into account second-order evidence, typically prior probabilities, where these are informative and thus epistemically relevant to the assessment of the viability of present evidence, is a failure of reasoning. To be more specific, it is a failure to consider objective – i.e., second-order - as well as subjective – i.e., first-order – factors, both of which are epistemically valuable. Otherwise put, to only consider first-order evidence and not to consider second-order evidence is a failure to consider *all* the relevant evidence.

This insight has been expressed in the literature. Kelly, for instance, articulates it as follows: “what it is reasonable to believe about the world on the basis of one’s evidence is constrained by what it is reasonable to believe *about* one’s evidence” (Kelly 2009, sec. 5.3). So too, Christensen notes that “the rationality of first-order beliefs cannot in general be divorced from the rationality of certain second-order beliefs that bear on the epistemic status of those first-order beliefs” (Christensen 2008, 18). To apply this insight to our previous example, this would mean that if, after encountering the Candlestick in the Hall I infer that Colonel Mustard did it, without considering that second-order evidence suggests that my inference that Colonel Mustard did it is 70% likely to be wrong, I would be neglecting relevant and thus epistemically valuable evidence.

2.7 THEORETICAL AND PRACTICAL PRESCRIPTIONS

The prescription to consider second-order evidence in conjunction with first-order evidence appears to be unproblematic in theory. In situations of uncertainty, second-order

⁴¹ See my paper on practical applications of Bayes’ rule.

evidence – usually prior probabilities – should be weighed against subjective likelihoods derived from first-order evidence, typically in accordance with Bayes’ rule⁴². But while in theory this appears unproblematic and straightforward, there seems to be a difficulty in realizing this prescription in practice.

Because the normative prescription is that second-order evidence ought to be incorporated in judgment in situations of uncertainty, it is subjective judgment that is responsible for determining a situation *as* uncertain, and consequently for determining *whether* second-order evidence is relevant as a risk-mitigating measure for addressing this uncertainty. But because determining whether a situation is uncertain depends, at least in part, on how confident the person is about first-order evidence, it can make the subjective judgment about *whether* second-order evidence is relevant dependant on the very unlawful outcome it is there to mitigate. And this is what paves the way to the real problem that a person that encounters first-order evidence appears to face.

The reason this is a practical rather than a theoretical problem is because in theory there is no apparent difficulty of incorporating second-order evidence in judgments about first-order evidence; in these kinds of theoretical cases, any judgment based on first-order evidence will be weighed against the second-order evidence that applies to that judgment.

⁴² Bayes’s rule, or theorem, is a rule for operating on numerically expressed probabilities to revise a prior probability (in other words, the base-rate) into a posterior probability after new data have been observed. According to the theorem, the posterior probability for event H1 *after* data D is observed and accounted for is: $p(H1|D) = p(H1) p(D|H1)/p(D)$, where $p(H1)$ is the prior probability assigned to H1 *before* D is observed.

Why? Well because this is the normative thing to do; second-order evidence is the means by which an individual judgment is put in the context of passed judgments and thus undergoes statistical corroboration. It is what is needed so that the viability of an individual judgment can be properly assessed. But in practice, whether or not second-order evidence applies to an individual judgment depends on how uncertain the person is or how compelling first-order evidence appears to be. If first-order evidence is compelling, second-order evidence may be taken to be irrelevant or inapplicable to that judgment because statistical corroboration appears unnecessary or inappropriate. Hence second-order evidence can be regarded as more or less applicable depending on the diagnostic value of individual judgment, which it was initially aimed to corroborate.

The normative problem that I am trying to outline is a problem of disagreement between peers considered from a practical point of view. And it is importantly different to theoretical abstractions of such disagreements. In theory, if you and I are peers and we disagree (even though each of us is confident about what we believe), second-order evidence about our peerhood is the means by which we locate ourselves in a statistical framework. The fact that we may find first-order evidence compelling seems to be neither here nor there; because disagreement exists between the two of us, and we are equally competent reasoners, second-order evidence must be used, since second-order evidence will have to be judged relevant to the case at hand in some sense and to some degree. In this case the question that remains is how to deal with first-order and second-order evidence in relation to each other. And this is what has traditionally been debated in the literature. But let us note that in practice, if you regard first-order evidence as sufficiently compelling, you will probably not regard the evidential situation as uncertain. And as

such, the corroboration of your belief with second-order evidence may seem irrelevant or even damaging to your subjective evaluation. In such a case, the fact that your peer disagrees may be evidence in favor of their being wrong. But it is important to see that even if, in such a case, you do not regard your situation as uncertain this does not mean that you are denying that there is a case of peer disagreement. You're not. Peer disagreement can still be a problem even if you yourself are confident about the first-order evidence. I mention this here because one might object that if you find that there is no uncertainty, then this is in fact a denial that there is a case of peer disagreement. But this only seems to be the case if one assumes that an agent's uncertainty about first-order evidence alone merits his using second-order evidence to corroborate his belief.

2.8 PRACTICAL PROBLEMS WITH THEORETICAL PRESCRIPTIONS IN DISAGREEMENT PROBLEMS

The practical difficulty with implementing the normative prescription to weigh first-order evidence against second-order evidence in situations of uncertainty is that from the first-person standpoint second-order evidence often has ambiguous implications. To see this, consider a hypothetical situation. Assume that I know that based on past performance I am 70% likely to be correct in my predictions. On first thought, I can take this to mean that there is a 70% chance that my next prediction will be correct and a 30% chance that it won't. Now, having made my next prediction there appears to be no way for me to ascertain, independent of relying on my present judgment and the various considerations that support it, whether my prediction falls in the positive or negative percentiles of chance. That is to say, I have no way of knowing whether my present belief is an instance affirming the 70% chance that I am correct or the 30% that I am not. My probability of

being correct, based on past performance, is second-order evidence the inclusion of which appears to depend on my present level of confidence about first-order evidence. And if I my confidence is high, I may regard my present judgment as being an affirming instance of the positive likelihood of my being correct (in line with how second-order evidence can be understood), rather than an uncertainty in need of corroboration. And because, on this interpretation, *I am* correct, there is no need to weigh my present level of confidence against second-order evidence⁴³. Note that I am not suggesting that I would necessarily be justified in neglecting the 70:30 base rate. Rather, what I am suggesting that the significance of the 70:30 base rate can mean different things in terms of epistemic justification.

This suggests that aside from the normative prescription to weigh first-order evidence against second-order evidence I appear to also have a normative obligation to treat first-order evidence in accordance with the level of epistemic warrant that it provides. It does after all seem to be uncontroversial that different evidential situations warrant varying levels of confidence according to their epistemic merit. And it seems that proper incorporation of new information about these situations depends on the epistemic warrant that is provided by first-order evidence. A person in a first-person standpoint may consequently be faced with two, possibly conflicting, normative prescriptions, from which the suspected practical paradox arises:

⁴³ Kelly states of cases such as these that “one’s first-order evidence not only confirms the belief in question; it also confirms a proposition to the effect that it is reasonable for one to hold that belief” (Kelly 2009, sec. 5.3).

- (1) Respond to first-order evidence in accordance with how convincing it appears to be.
- (2) Mitigate the risk of being wrong by weighing first-order evidence against second-order evidence.

As noted, the practical problem here seems to be that there is no independent way to ascertain which prescription – (1) or (2) – applies, and there is no immediately obvious or straightforward way of combining them. And this situation, in which we are asked about an individual's appropriate response once conflict with a peer is discovered, appears to be reflective of a class of epistemically ambiguous situations in which the crux seems to lie in an individual's ability to determine the appropriate revisionary response to the situation. And because doing so is largely a matter of how confident the individual is about the first-order evidence, the inclusion of second-order evidence seems to depend, at least in part, on the selfsame risk-prone reasoning it is there to mitigate. And this, it seems, is inevitable. As long as subjective judgment is responsible for deciding *whether* second-order evidence is relevant, it doesn't appear to matter that second-order evidence is independent of current judgment, the risk of fallible subjective judgments continues.

2.9 THE REAL PROBLEM OF DISAGREEMENT – A *PRACTICAL* PARADOX

If the normative prescription is that an individual in a decision circumstance ought to decide whether or not second-order evidence is relevant to that circumstance, he must have the ability to distinguish when it is and when it isn't relevant. We assume that this is a function of how ambiguous the evidence is, perhaps how weak the evidence is (Enoch 2010; Kelly 2005), or how uncertain he is about what he has inferred from the evidence.

In each case it is on the basis of such considerations that the individual deems second-order evidence relevant or irrelevant. And this seems to lead to a situation where the judgment of relevance has no independent evidence to go by. Hence theoretically, Bayes rule may well offer a precise way to weigh beliefs, and philosophers writing about disagreement may well make suggestions about which responses are appropriate for peers that encounter conflicting beliefs. But tactics such as these only seem to be relevant *after* the hard work of deciding relevancy has been done, and this, as I have tried to show, is theoretically an underdescribed problem that I believe has largely been missed until now⁴⁴. It also captures what I take to be the *real* problem of disagreement.

3 RESPONSES IN THE LITERATURE

I believe that responses to peer disagreement in the recent philosophical literature can be divided into three kinds. In the next section I will address each of the approaches in the literature in relation to the problem highlighted in section 2. Before doing so I will sketch a pseudo-particularized example of disagreement on the basis of which the plausibility of each of the approaches can be assessed.

3.1 A CASE OF PEER DISAGREEMENT

⁴⁴ Elsewhere I discuss the epistemic significance of relevancy judgments. In this context see also Maya Bar-Hillel's seminal "The base rate fallacy in probability judgments" (Bar-Hillel 1980), which focuses on relevancy judgments in establishing whether or not base rates ought to be incorporated in probability judgments. See also: (Bar-Hillel and Fischhoff 1981; Bar-Hillel 1982; Welsh and Navarro; Barbey and Sloman 2007; Ajzen 1977).

Jill and Jack are two equally ranked chess masters. As it happens, Jill and Jack have other things in common aside from sharing the same title and rank at chess. They have both been playing chess for the same number of years and they have won the same number of games, at equally ranked tournaments, against equally classed players, using similar game strategies. Additionally, Jill and Jack also know all of this about each other.

On a particular occasion, Jill and Jack are each independently asked by an examiner which color has the advantage in a particular chess-board arrangement. Jill tells the examiner that she thinks that White has the advantage; Jack tells the examiner that he thinks that Black has the advantage. Then each of them is told by the examiner about what the other thinks. What should Jill and Jack do in regard to their beliefs after being given this information, assuming that is, that neither one of them has any non-question-begging reason to think that the other happens to be reasoning in sub-standard conditions – that, for instance, the other isn't drunk, dazed, tired, or anything of the sort? More specifically, should the discovery that they believe differently make either of them lose confidence in their own beliefs?

3.2 THE BOOTSTRAP APPROACH⁴⁵

The bootstrap approach with regards to disagreement makes use of a person's own reasoning about the issue at hand to support his revised reasoning about the issue at hand. It suggests that because *P* is true it doesn't matter that a peer disagrees about this, *because P is true*, and one's peer is therefore wrong. This is blatantly question-begging

⁴⁵ Elga (Elga 2007) also discusses the problem of bootstrapping, as do Kelly (Kelly 2009) and Enoch (Enoch 2010). I do not discuss either of these views regarding bootstrapping here.

reasoning. And it is reasoning that fails to appreciate the epistemic significance of disagreement⁴⁶. Moreover, for disagreement to have no epistemic impact is in fact a failure to appreciate it as relevant evidence⁴⁷.

The bootstrap response ignores second-order evidence concerning peerhood and contends that the appropriate response to the discovery of disagreement is to act in accordance with what first-order evidence suggests⁴⁸. This would mean that Jill would take her belief that White has the advantage to support the belief that Jack is wrong *because* White has the advantage⁴⁹. She would thus *not* adjust her credence in White having the advantage

⁴⁶ This is why David Enoch has fittingly called it the “*I don’t care view*” (Enoch 2010, 15); the view is attributed to Thomas Kelly (Kelly 2005). Neither Enoch nor Kelly contend that this is a plausible response to disagreement, largely because it completely ignores the epistemic significance and corrective role that other people’s opinions can have on our own judgment.

⁴⁷ Matheson presents a novel argument for why evidence of disagreement is, after all, *relevant* evidence. According to Matheson, if, to continue our chess example, Jill were to ask Jack which color he thinks has the advantage on the present chess-board position, she would be justified in believing Jack on this matter, since she relies on Jack (in fact she relies on Jack on these matters as much as she relies on herself). If this is so then Jack’s belief *does* provide Jill with evidence that Black has the advantage. And it is therefore evidence for Jill regardless of what she believes (Matheson 2009).

⁴⁸ In doing so it violates what Christensen has recently called “independence”: “In evaluating the epistemic credentials of another’s expressed belief about P, in order to determine how (or whether) to modify my own belief about P, I should do so in a way that doesn’t rely on the reasoning behind my initial belief about P” (Christensen 2011).

⁴⁹ This is someone reminiscent of Kelly’s view according to which a disagreement is already an indication that asymmetry exists between peers; the fact that there is disagreement (and assuming epistemic permissiveness is not permitted) already indicates that someone is right and someone is wrong. And

because from her standpoint she has no reason to. Her disagreement with Jack appears to be both epistemically insignificant, *because he is wrong* and thus what he thinks is irrelevant, and misguiding, because he is *wrong*⁵⁰.

The bootstrap approach is normally referred to as a limit case expressing the tempting appeal of the kind of unwarranted reasoning that unregulated internal standards for justification can sometimes give rise to⁵¹. All the same, what characterizes the bootstrap approach is that it only lets subjective credence in relation to first-order evidence influence its judgment. Second-order evidence has no effect. The bootstrap approach is especially dangerous because the person using the bootstrap approach may well argue that his tactic *is* to use the outputs of first-order evidence *and* second-order evidence but

supposedly, the one who got it right should *not* revise his beliefs (Kelly 2005). See also Enoch for a critical evaluation of Kelly's early view (Enoch 2010).

⁵⁰ This line of thought is somewhat reminiscent of Kripke's dogmatism paradox. Gilbert Harman transmitted Kripke's Dogmatism Paradox (Harman 1973), which is also presented in revised form in (Kripke 2011). See also Kelly's treatment of the paradox in relation to disagreement and higher-order evidence (Kelly 2008).

⁵¹ It is, in this context, sometimes referred to as the Extra Weight View. See: (Elga 2007; Kelly 2009; Enoch 2010). In a wider context, the bootstrap approach disregards the awareness that people normally have of their own fallibility and the corrective role that they attribute to other people's opinions as a means for arriving at better conclusions. It grants first-person conviction about first-order evidence a justificatory role that is normally attributed to independent standards. In so doing it uses circular reasoning to bootstrap the justification of a belief to conviction about it. In itself, this doesn't necessarily pose a problem, yet once relevant second-order evidence is available, dismissing it altogether is wrong.

that occasionally the output of first-order evidence overrides the output of second-order evidence and thus makes it *irrelevant*.

What seems to be the core problem with bootstrapping is that second-order evidence is relevant after all, even in those cases where it is supposedly swamped by the epistemic force of first-order evidence. What makes it relevant is that the viability of any single judgment that belongs to a statistical data set of similar judgments ought to be established in reference to that data set, or, more specifically, it ought to be weighed against second-order evidence pertaining to it. Nonetheless, what is intuitively appealing about bootstrapping, is that whether a single judgment belongs to a particular data set will depend on the uncertainty with regards to the case at hand. If there is uncertainty, then corroboration is in order. If there isn't uncertainty, or if the degree of uncertainty is small, then this may merit exclusion from the data set to which such an instance would otherwise belong.

3.3 THE CONCILIATORY APPROACH

The conciliatory approach is probably the most intuitively reasonable approach because it appreciates the epistemic significance of disagreement and respects the corrective role that other people's beliefs can have. This approach recognizes both the normative requirement to respect the epistemic force of first-order evidence and the normative requirement to consider second-order evidence when things are uncertain. Accordingly, that Jack believes differently to Jill about what color has the advantage in the present alignment of the pieces on the board *is* epistemically significant evidence and should be incorporated into Jill's response once she finds out what Jack believes. In fact,

recognizing that Jack believes differently ought to make Jill doxastically shift in the direction of Jack's belief (Matheson 2009).

The conciliatory approach contends that Jill's confidence in White being advantageously positioned should be revised following her discovery of Jack's belief. But it does not state *how* it should be revised, or more specifically, to what degree⁵². The answer to this question seems to depend on how *relevant* Jill believes Jack's belief is for her revised response. It depends on the measure of epistemic significance that she grants it.

We shall return to the conciliatory approach momentarily, after examining the egalitarian approach. In the meantime let us note that the merit of the conciliatory approach appears to be that it recognizes that what an equally competent person believes *is* relevant evidence. Moreover, it prevents bootstrapped dismissals of what other people think. All the same it leaves open the question regarding the weight that ought to be granted to prior knowledge concerning peerhood. And on this matter the egalitarian approach, which is a particular kind of conciliation, provides an answer concerning the appropriate weight that ought to be granted to second-order evidence so as to reach rational epistemic compromise.

3.4 THE EGALITARIAN APPROACH

The egalitarian approach to disagreement says the following. If two people are equally likely to be correct and they unexpectedly discover that they hold different beliefs about what *is* correct then they ought to recede their confidence in their own belief being

⁵² Matheson (Matheson 2009) clarifies and defends weak and strong conciliatory responses, from very little movements to strong movements, entailing, for instance, the suspension of judgment.

correct to the conditional probability that would be granted to their being correct in considering, prior to the actual disagreement, the appropriate response were such a situation to occur⁵³. And because they are equally likely to be correct, the appropriate conditional probability that they ought to grant their own beliefs would be 0.5. Why 0.5? Because from the theoretical vantage point, the probability that they would be correct if such a situation would occur is equal. If there were more than two peers involved in such a scenario that the probability that any one of them is correct conditional on a disagreement occurring would be $1/n$. Otherwise put, if two peers disagree and neither has any reason aside from the other's opinion to think that the other is wrong, such as their possession of more or better evidence or their superior conditions for inferring from the evidence in the present circumstance, then the epistemic weight of first-order evidence ought to be ignored, and the revised credence ought to be based on second-order evidence alone.

The egalitarian approach suggests what Sarah Moss has called a “perfect compromise⁵⁴” (Moss 2012). If A assigns credence C_1 to P, and B assigns credence C_2 to P, then a ‘perfect compromise’ would be for A and B to assign $(C_1 + C_2)/2$ credence to P (Moss

⁵³ This is largely based on Elga's formulation for the Equal Weight View (Elga 2007): “Upon finding out that an advisor disagrees, your probability that you are right should equal your prior conditional probability that you would be right. Prior to what? Prior to your thinking through the disputed issue, and finding out what the advisor thinks of it. Conditional on what? On whatever you have learned about the circumstances of the disagreement” (Elga 2007).

⁵⁴ Adam Elga's (Elga 2007) and David Christensen's (Christensen 2007) views are two examples of what I take to be the egalitarian approach to peer disagreement. On both of these accounts, if disagreement is apparent, each peer ought to revise their confidence in alignment with what second-order evidence dictates.

2012). It is important to see that according to Moss' position, the credence value of C_1 and that of C_2 are given equal weight – e.g., each of them is given 0.5, as above.

The egalitarian approach amasses its conclusions from theoretical thinking about a *possible* disagreement occurring. It contends that if two people know that based on their prior performance and capabilities they share an equal probability of being correct on some hypothetical occasion, and they also consider (from this same hypothetical perspective) the probability of their each being correct on the occasion of such a hypothetical disagreement occurring, then their *actual* response ought to be equal to their hypothetical response, that is, treating each belief as equally probable.

Aside from the fact that egalitarianism completely swamps first-order evidence with second-order evidence, it doesn't take into account that in an *actual* case of disagreement each of the parties involved has reasons for believing as they do and possibly also epistemic grounds for dismissing the purely hypothetical revisionary suggestion. The actual disagreement as opposed to the hypothetical disagreement appears to possess an epistemic factor that the theoretical consideration neglects to consider. In the actual disagreement one person may have reason to think that they are correct in light of the epistemic force of first-order evidence, and as a consequence to think that the other person is incorrect⁵⁵. From the theoretical vantage point this kind of playing of subjective

⁵⁵ And in the actual disagreement this might be explained in a number of ways. The dissenting person, in view of his divergence of opinion, may be thought to have slipped in performance - made a mistake that is, a performance-error, perhaps misapplying the proper rules of inference. Alternatively, the divergence itself may be regarded as a reason, or perhaps even a proof, that that person ought to be demoted from the level of peer.

credence against second-order evidence is unwarranted. But from the actual perspective of a person in the midst of disagreement the justificatory weight of subjective credence is not only plausible but also appears to be a normative requirement – a requirement to respond to the evidence in a manner that is sanctioned by *that* evidence (prescription #1 above)⁵⁶.

The merit of egalitarian suggestions for belief revision is that they obey prior probabilities in the absence of any non-question-begging circumstantial indicators. Their limitation is that they eradicate the epistemic weight of the first-order considerations on the basis of which Jill came to believe as she did in the first place. Egalitarianism contends that if there is to be compromise the only compromise there can be is perfect compromise, while actual disagreements suggest that it isn't at all clear that this is so.

There is a danger in the egalitarian approach. It is the danger of rigid reasoning. If a person does not honor the prior probabilities entailed by second-order evidence but instead defers to them completely, his decisions and behavior will be more rigid; new experiences will be classified on the basis of previously established probabilities and information will be absorbed less for its intrinsic value and more according to whether it meets the terms of a bet. Hence the question becomes how the diagnostic value afforded by first-order evidence can be respected while the dictates of second-order evidence are honored too.

⁵⁶ To dismiss first-order evidence altogether, as egalitarian positions seek to do, leads to a sort of real-world skepticism (Feldman 2006, 415), according to which the level of credence attributed to many of our commonly held beliefs ought to be reduced, which is implausible.

3.5 BACK TO THE CONCILIATORY APPROACH

Once we recognize that actual disagreements involve reasoning on the basis of everything that is epistemically available to us (Enoch 2010, 38), including subjective contentions regarding the conclusiveness of first-order evidence, the dangers of bootstrapping seem to be inevitable. All the same, to completely overrule first-order evidence on the basis of second-order evidence, as the egalitarian approach suggests, seems wrong. In this sense the dangers of bootstrapping are dangers that we will perhaps have to live with (Enoch 2010). But without preserving an individual's normative obligation to deal with these dangers we would seem to have to give up too much of what good reasoning amounts to and what plausible revisionary approaches to disagreement ought to offer. Indeed, to eradicate the possibility of error altogether by dismissing any possibility of inaccuracy would seem to compromise our ability to encounter and conditionalize on new evidence.

The conciliatory approach seems to provide the most plausible tactic for responding to disagreement. But because it doesn't specify a specific rule for how to balance first-order evidence with second-order evidence, it appears to leave much of the hard work to the individual's sense of judgment. In doing so it is perhaps less explicit in its rules for revision but it is, nonetheless, more attuned to the complexity involved.

4. KELLY'S TOTAL EVIDENCE VIEW

The Total Evidence View proposed by Thomas Kelly (Kelly 2009) seems to recognize some of the concerns that I have raised here. Kelly rightly contends that upon encountering disagreement with a peer, the interaction between first-order and second-

order evidence can be a complicated affair. And that in such cases what it is right to believe “depends on both the original evidence as well as on the higher-order evidence that is afforded by the fact that one’s peers believe as they do” (Kelly, in: Whitcomb 2011, 201). The merit of Kelly’s Total Evidence View appears to be that it acknowledges that there is a difference between theoretical problems of disagreement and actual problems of disagreement, a difference that I have focused on here. But unfortunately Kelly doesn’t peruse the normative significance of these observations further⁵⁷. There are lines of similarity between what I am proposing here to be the real problem of disagreement and Kelly’s Total Evidence View. I do not mean for my view and Kelly’s to be exclusive in any way. As noted, I think that Kelly recognizes the difficulty posed by egalitarian solutions and he also notes the importance of psychological considerations relating to first-order evidence⁵⁸. All the same, what I have proposed here takes Kelly’s observations one step further.

Given the importance of incorporating first-order evidence as well as second-order evidence, and seeing as psychological persuasion about first-order evidence can

⁵⁷ Kelly has recently pursued this direction further in his, “Disagreement and the Burdens of Judgment,” unpublished but available online at: <http://www.princeton.edu/~tkelly/datbj.pdf>.

⁵⁸ Kelly also observes that a person’s psychological persuasion about first-order evidence may determine the way in which they consider second-order evidence: “In some cases, the first order evidence might be extremely substantial compared to the higher-order evidence; in such cases, the former tends to swamp the latter. In other cases, the first order evidence might be quite insubstantial compared to the higher order evidence; in such cases, the latter tends to swamp the former” (Kelly, in: Whitcomb 2011, 202). This appears to be an important observation about human reasoning and in particular it emphasizes the force that psychological conviction can have on the assessment and evaluation of evidence.

sometime swamp the weight of second-order evidence, whether or not it *will* swamp second-order evidence appears to depend on the subjective point of view. In effect this means that whether or not second-order evidence will be weighed or downgraded, and to what extent, will depend on what is deemed appropriate, on grounds of relevancy, by subjective appraisal. And therefore, as I have tried to argue, the real problem of disagreement is that one must respect the epistemic force afforded by first-order evidence while honoring the dictates of second-order evidence. And this inevitably leads to a situation where the epistemic force of first-order evidence determines whether or not second-order evidence is relevant. Revising beliefs without attending to first-order evidence seems implausible. But letting conviction about first-order evidence determine whether or not second-order evidence is relevant paves the way to the error prone reasoning that revisionary tactics are there to mitigate.

And in recognizing this core problem, perhaps Kelly's Total Evidence View is not a revisionary tactic for dealing with disagreement at all, but rather an acknowledgment of I) the need to weigh first-order as well as second-order evidence, and II) an appreciation of the role that psychology has in determining whether or not second-order evidence should be taken into account. My account of the real problem of disagreement and the emphasis on practical difficulties with the subjective appreciation of second-order evidence takes disagreement to point to a more general phenomenon of practical reasoning, in which relevancy judgments – such as determining whether second-order evidence is appropriate and should be applied – form a central part.

5. SO WHAT, AFTER ALL, CAN WE DO?

As we saw, subjective judgments relating to second-order evidence can be dangerous because they may be based on circular reasoning. All the same, to avoid these dangers by dismissing the dictates of subjective judgment altogether upon encountering disagreement is implausible because it will entail that many beliefs about which we are perfectly confident will have to be dismissed. And it will also undermine the ability to attain new information whenever conflict is encountered. The appropriate response to disagreement ought to honor both the normative requirement to consider second-order evidence and to respect first-order evidence in proportion to the epistemic force that it provides. How these two requirements should be considered in an actual case of belief revision – how much weight ought to be granted to each - is something that must inexorably be left to the judgment of the subject that is aware of these two requirements. And this, as I said at the outset, is inevitable.

What I have tried to do here is to highlight a neglected problem distinct from the problem of disagreement as it is usually discussed. This problem stems from an intrinsic property of second-order evidence that can make otherwise clear data, typically prior probabilities, practically ambiguous. In actual situations of disagreement the problem of determining how relevant second-order evidence is appears to be more fundamental than establishing the rational way of combining first and second-order evidence. And incorporating second-order evidence often depends on determining whether it is epistemically appropriate, and this, as I have tried to show, will often be a function of the subjective appreciation of the epistemic force of first-order evidence.

As a concluding remark, and so as to make sure that the problem I have raised here is not mistaken as being merely another position in the existing debate, let me point to why the present observation is novel. The problem of relevancy judgments and the ambiguous practical implications of second-order evidence suggests that even if a particular revisionary response is considered rational in theory – suggesting for instance such-and-such a combination of first and second-order evidence – in practice it may be deemed inapplicable on grounds of the epistemic force, or diagnostic value, of subjective judgments about first-order evidence.

Abstract: in this paper I present a criticism of Sarah Moss' recent proposal to use scoring rules as a means of reaching epistemic compromise in disagreements between epistemic peers that have encountered conflict. The problem I have with Moss' proposal is twofold. Firstly, it appears to involve a double counting of epistemic value. Secondly, it isn't clear whether the notion of epistemic value that Moss appeals to actually involves the type of value that would be acceptable and unproblematic to regard as epistemic.

INTRODUCTION

In a recent paper, Sarah Moss has suggested (Moss 2012)⁶⁰ that when they encounter conflict, epistemic peers should not split the difference between the credence that they each assign to some disputed proposition P, as has been suggested by conciliatory approaches to belief revision in the debate surrounding disagreement in the literature (Elga 2007; Christensen 2007; Matheson 2009). Moss contends that an epistemic compromise between epistemic peers need not be the arithmetic mean of prior credences (Moss 2012, 2); if my credence in some proposition P is x and yours is y, the credence that is the result of our compromise need not be $(x + y)/2$, as an equal weighing of our beliefs would suggest. Instead, Moss proposes an alternative strategy for how epistemic peers *ought* to compromise. The basic idea is that splitting the difference (Christensen 2007) between credences may be inadequate seeing as agents may assign different

⁵⁹ For a published version of this chapter in *Episteme* see: (Konigsberg *forthcoming*).

⁶⁰ Moss, Sarah (2012). "Scoring Rules and Epistemic Compromise." *Mind* 120 (480): 1053-1069.

epistemic values to different credences. It is because of this that a compromise that would only consider credences would lack the *epistemic value* that those credences have for the agents involved.

These differences between how different agents evaluate credences supposedly make for *real* differences with regards to what compromising agents *should* do (Moss 2012, 8). It is on these grounds that Moss contends that a more plausible suggestion for epistemic compromise is that disagreeing agents should maximize their *expected epistemic value*. Roughly, they should maximize the epistemic value that alternative credences have for them. This, in a nutshell, is Moss' suggestion for proper epistemic compromise.

While novel and clearly argued, I think that Moss' proposal fails to provide entirely convincing reasons for abandoning the traditional symmetrical approach to epistemic compromise (e.g., Elga 2007; Christensen 2007) and for adopting the scoring rule model instead. In what follows I demonstrate two problems with the model that Moss advocates. The paper is composed of two sections. In the first section I present and discuss Moss' proposal in context. In the second part of the paper I present and expound on two problems on account of which I find Moss' scoring rules solution for epistemic compromise inadequate.

I

SCORING RULES

One of the most important factors to consider in assessing normative tactics of belief revision and compromise is the way in which decision makers quantify uncertainty. In

many cases related to in the literature, decision makers quantify uncertainty by expressing their own beliefs using probabilities. In some cases, they don't express their own beliefs but instead defer to the belief of a counterpart or a source of information who they feel is more capable of making such judgments, or has some information that they don't have. In situations such as these decision makers use their counterpart's beliefs as a model for what to believe (J. M. Joyce 2007). But in both cases – whether they express their beliefs using probabilities or decide instead to defer to someone else's belief – their choice is rooted in personal judgment. And personal judgment, as we all know, is not infallible. Because it is not infallible, various methods have been devised by which decision makers can improve their assessments. One such method is the scoring rule.

Traditionally (e.g., Brier 1950), scoring rules measure the inaccuracy of a set of probability assessments⁶¹. They do so by relating an agent or an entity's expected performance to their actual performance through calibration. From an *ex post* point of view, scoring rules can provide a means by which agents can be evaluated with respect to their predictive abilities (Jose 2009, 264). Conceived of in this way, scoring rules enable decision makers to adjust and improve their predictive assessments based on calibration of past expectations versus the past performances to which those expectations relate. If, for instance, an agent's past performance shows a 0.3 difference between actual performance from expected performance – i.e., as expressed by credences assigned

⁶¹ For recent work on using scoring rules see: (Ehm 2012; Werner Ehm and Gneiting 2009; Gneiting and Raftery 2007; Jose 2009; Bickel 2007). In specific reference to the use of scoring rules to assess credences see: Gibbard, "Rational Credence and the Value of Truth" in: Gendler and Hawthorne 2008, chap. 6; Percival 2002).

versus performance, then this function will comprise the scoring rule by which present predictions, or levels of confidence, *should* be adjusted. And by adjusting credences, scoring rules can progressively refine margins of error and achieve greater predictive accuracy.

In her novel contribution to the debate surrounding rational responses to disagreement, Moss suggests that epistemic compromise between disagreeing parties can be achieved using scoring rules. But the scoring rules that Moss has in mind are different to traditional scoring rules. Scoring rules as Moss conceives of them are functions that, roughly speaking, measure epistemic value. They are functions that describe the epistemic value, in the agent's eyes, of having credence x in some proposition p , in cases where p ends up being true and in cases where p ends up being false. In the sense in which Moss uses them, scoring rules are functions that measure positive as well as negative epistemic value. In making use of scoring rules construed in this way, Moss is following others that adopt this interpretation (e.g., Gibbard 2007, Joyce 2007; Joyce 1998; Joyce 2009; Oddie 1997; Percival 2002) and seek to construe epistemology decision-theoretically (the value to be maximized by beliefs or credences being truth-related). I will elaborate more on the notion of epistemic value shortly. For the meantime the important point to note is that Moss contends that “using scoring rules ... agents may compromise by coordinating on the credences that they collectively most prefer, given their epistemic values” (Moss 2012, 2). The fundamental idea advocated is that instead of compromise being conceived of as the arithmetic mean of prior credences, it should be understood as the mean that maximizes expected epistemic value.

EPISTEMIC VALUE

Moss' discussion of epistemic compromise is located in the wider context of epistemic rationality, in which purely epistemic considerations are needed for the desired form of compromise to be rational⁶². The problem that Moss has with conciliatory approaches to compromise is that they do not take into account the fact that agents may value some credences more than others, in what she regards as a "purely epistemic sense." These approaches only look at credences, whereas they should also be looking at the epistemic value that different credences may have. And even though Moss departs from these positions, this is not because she believes that they violate the conditions of epistemic rationality by, for instance, considering *non*-epistemic factors in their proposed tactics for epistemic compromise. She departs from them because she believes that they don't consider *all* the epistemic factors that should be considered in epistemic compromise.

If it is to suggest how agents ought to compromise in a purely epistemic sense, the solution that Moss proposes for rational epistemic compromise must also incorporate only epistemic factors. Moss believes that epistemic value is the component that is lacking in conciliatory approaches. And she also believes that its incorporation is what makes her solution to compromise preferable. Perhaps more importantly, Moss contends that epistemic value is a purely epistemic kind of value. And it is that which consequently allows her proposal for epistemic compromise to remain in the domain of epistemic rationality. Following Gibbard (Gibbard 2007), Moss regards epistemic value as the

⁶² For more on the notion of epistemic rationality see:(Armstrong 1973; Harman 1995; Harman 1997; Kornblith 1993; Kelly 2003; Kelly 2002; Foley 1987; Foley 1993; Feldman 2000; Wedgwood 2002; Velleman 2000; Nozick 1994; David 2001).

preference that an agent has for a particular credence in relation to one of the contested goals of epistemic rationality, roughly, that of believing true things. The kind of preference she seems to have in mind is that according to which an agent may care more about some types of credences than about others, on purely epistemic grounds.

CARING ABOUT CREDENCES

If all agents had monotonic scoring rule functions, they would agree that greater value should be placed on more accurate estimates. Thus if there were no divergence in attributions of epistemic value between agents in cases where the estimates in question are estimates of the truth value of true propositions, estimates that are closer to 1 would always be better than those that are closer to 0, and a credence such as 0.8 would be more epistemically valuable than credence 0.7, because of its proximity to certainty. So too, in this case the difference between credences would be valued similarly by different agents: the *difference* between 0.7 and 0.8 would have the same epistemic value as that between 0.2 and 0.3, and as any other pair of credences differing by 0.1. And this would be true for all agents in all situations. But apparently, agents may not share a uniform evaluative scoring rule for credences, even in purely epistemic terms. In fact, believing true things may mean different things for different agents, depending on the credences which they assign.

For instance, we can envision that the function f which describes A's scoring rule of P's possible credences may be expressed as: $(f_1(0.9) - f_1(0.8)) > (f_1(0.6) - f_1(0.5))$, which means that the difference in epistemic value, for A, between 0.9 and 0.8 is greater – it is preferred to – the difference between 0.6 and 0.5, even though the arithmetic difference

between these expressions is the same. A possible reason for this might be that for A, the former difference in credences relates to estimates of truth that are closer to being certain than the latter; e.g., $[1 - (f_1(0.9) - f_1(0.8))] > [1 - (f_1(0.6) - f_1(0.5))]$. And it is the latter which A prefers. But we might also imagine another agent – B, whose scoring rule of P's credences is expressed as: $(f_1(0.6) - f_1(0.5)) > (f_1(0.9) - f_1(0.8))$, where the arithmetic difference between lower credences is more epistemically valuable for B than the same arithmetic difference between higher credences that have a greater positive proximity to certainty. In this sense for A to assign credence 0.7 to P is different than for B to do so. Meaning that in such a case, the epistemic value of 0.7 for A and for B may be different.

It is because the epistemic value of credences may vary between agents that strategies of compromise between agents ought to take this into account. Not to take this variation in preferences into account would not only be to consider insufficient factors – i.e., only considering credences and not considering their epistemic value, but also to consider misleading factors – because credences on their own, on Moss' conception, do not express the epistemic value which they carry.

Because different agents have different scoring rules by which they assign values to credences, a simple, purely arithmetic compromise between the credences of different agents will not constitute a perfect compromise of opinion. Seeing as the scoring rules which assign value to credences can vary between agents, a merely arithmetic average of the disputed credences may not amount to what we would expect from a perfect epistemic compromise, contra to what some conciliatory approaches would suggest.

Hence if compromise means a settlement of differences or mutual concessions, and expressions of epistemic differences between agents are not only credences but also the epistemic values attributed to credences, then a proper epistemic compromise should include both of these factors. Moreover, it is also plausible to assume that a satisfactory mutual concession between disagreeing parties would be that which would be preferred by both parties. And that which would be preferred would seemingly be that which maximized expected epistemic value.

Accordingly, Moss proposes that the “perfect” epistemic compromise will be the maximization of the average epistemic value of the credence of a proposition for each of the agents in question. It is, she suggests, the coordination of the agents “on the credences that they collectively most prefer, given their epistemic values” (Moss 2012, 2).

II

WHY THE SCORING RULE SOLUTION MAKE SENSE

Moss seems to take her scoring rule solution to epistemic compromise to be justified for the following reasons⁶³:

- The scoring rules solution is intuitively reasonable;
- Because it averages combined epistemic values, the scoring rule solution ensures that each party in the dispute is given *something* – what we might regard as a fair share - that they intrinsically care about;

⁶³ I concur with the three factors that (Shultz) identifies that Moss appeals to. See: (Shultz).

- The scoring rules solution ensures that neither party in the dispute gets swamped by the epistemic values that another party grants to the same credences;
- The process of compromising by maximizing expected epistemic value never recommends a credence such that there is some other particular credence on which all agents would *prefer* to compromise;
- Compromising by maximizing average expected epistemic value is analogous to an intuitively reasonable method of practical compromise in which the average expected practical value of an outcome is maximized⁶⁴.

Having presented Moss' solution and its advocated benefits, I will now proceed to examine how feasible it really is.

THE FEASIBILITY OF THE SCORING RULES PROPOSAL

Moss' proposal is novel, intuitively appealing, lucid, and coherent. The paper discusses important distinctions that are worthy of attention in their own right. These include credence-eliciting and non-credence-eliciting scoring rules (Moss 2012, 7–9)⁶⁵. But setting the merits of the paper aside, my present concern is with the viability of the model that Moss proposes. More specifically, with the contention that scoring rules provide an understanding of “how an agent may value certain credences over others, *in a purely*

⁶⁴ I am grateful to an anonymous reviewer for drawing my attention to points #4 and #5.

⁶⁵ See also: (Gibbard, Allan, “Rational credence and the value of truth.” 2006. In: Tamar Szabo Gendler and Hawthorne 2008).

epistemic sense” (Moss 2012, 2; my emphasis)⁶⁶, as well as with the way in which Moss takes this idea to apply as a preferred solution for epistemic compromise.

To put things concisely, the problem I have with Moss’ proposal is twofold. Firstly, it appears to involve what I would like to refer to as a double counting of epistemic value. And secondly, it isn’t clear whether the notion of epistemic value which Moss appeals to actually involves the type of value that it would be acceptable and unproblematic to regard as epistemic. And on account of these two concerns, I believe that Moss’ proposal fails to provide entirely convincing reasons for abandoning the traditional symmetrical approach to epistemic compromise (Elga 2007; Christensen 2007) and for adopting the scoring rule model instead.

That said, I am not sure that symmetrical approaches such as splitting the difference between credences (Christensen 2007) or granting credences assigned by disagreeing agents equal weight (Elga 2007) are the perfect way to compromise epistemically. But nevertheless, neither do I think that Moss’ proposal is sufficiently convincing. And this is what I will focus on demonstrating presently. I shall now set forth the two problems I see in the scoring rules approach to epistemic compromise that Moss advocates.

DOUBLE-COUNTING

Credences are truth-value estimates (J. M. Joyce 1998), and these estimates are subjective estimates. They are an agent’s evaluation of the probability that some proposition is true

⁶⁶ In this Moss is not alone, and my argument applies to others that hold this view. Nonetheless, my focus in this paper is on Moss’ account.

or an agent's evaluation of the level of confidence that a belief in such a proposition being true deserves. Because of the partially subjective origins of truth-value estimations, the way in which people value credences, insofar as this value is epistemically significant, should seemingly *already* be incorporated into the process according to which credences are assigned⁶⁷.

It seems plausible to assume that an estimation of the truth value of p on the basis of evidence E is at least partly a matter of the degree to which p seems to be *supported* by E. And the notion of 'support' here entails at least some degree of subjective appreciation of the credence that ought to be assigned to p in light of the evidence. And this would be in addition to the assumption that the credence assigned to p is, or can also be, partly directly responsive to E, above and beyond the way E is subjectively understood.

Hence assigning credence is not something that is wholly a function of the available evidence; it is also a function of what that evidence is taken to mean, and the degree of support it is taken to afford the proposition in question. As such, the confidence that the evidence affords the proposition in question is what leads an agent to assign it a particular credence. And the credences chosen already incorporate the agent's preferences regarding the epistemic value of the range of credences for the proposition in question. Otherwise put, the credence chosen as an assessment of p's viability, or of how much

⁶⁷ It seems worth mentioning that if your scoring rule is credence-eliciting – which it arguably must be if you are rational – then incorporating your epistemic value into your assignment of credences will not change which credences you assign to propositions, since adjusting your credences to maximize their expected epistemic value will leave your actual credence alone. I discuss this point later in the paper. I am grateful to an anonymous reviewer for this point.

confidence p deserves, derives from the function that expresses epistemic value across a range of credences.

If this is indeed the case, then counting the epistemic value of credences separately from the credences themselves appears both superfluous and distortive for reaching epistemic compromise. In this sense a credence that has been assigned to p does not appear to be distinct from the scoring rule according to which that credence is epistemically evaluated. Rather, p 's credence seems to be the *result* of the scoring rule by which it is determined. Let me explain.

If these were traditional scoring rules, for instance a Brier score or a Spherical score, and p 's credence was 0.7, the scoring rule would not appear to modify the credence assigned to p . In this case, one could switch the scoring rule for p from f_1 to f_2 , for instance from a Brier score to a Spherical score, and the credences assigned to p would not change, it would remain 0.7⁶⁸. The scores would, in both these cases, calibrate predictions with regard to p , rather than alter these predictions.

But as Moss relates to them, scoring rules do not measure the performance of predictions made under uncertainty. They instead measure the epistemic value of assigning a range of credences to p (See section I, *above*); as Moss regards them, scoring rules are functions that describe the epistemic value, in the agent's eyes, of having credence X in some proposition p , in cases where p ends up being true and in cases where p ends up being false.

⁶⁸ I am grateful to an anonymous reviewer for this point.

Hence on this understanding of the role of scoring rules, the credence that an agent assigns to p should be taken as an expression of the scoring rule used for assigning epistemic value to p . If, for instance, the agent happens to be indifferent to small differences between credences – say, the difference between 0.7 and 0.8 – as long as both credences are in the right direction – e.g., with proximity to truth – i.e., 1 - not falsehood – i.e., 0, then this should already be apparent in the credence that has been assigned to p – i.e., whether it has been assigned credence 0.7 or 0.8⁶⁹. And in the same way, if one is indifferent to the difference between credence 0.2 and 0.3 when it comes to the likelihood of falsehoods – i.e., that one is indifferent to the difference between 0.2 and 0.3 when these are assigned to p , if p ends up not being true, then this should *already* be apparent in the credence that one has previously chosen to assign. And therefore considering epistemic value *after* it has served its purpose is both superfluous and distortive. The role granted to scoring rules on Moss’ account suggests that the credence assigned to p by A , derives from the scoring rule f , which expresses the epistemic value of that credence in relation to p for A .

It is worth mentioning that at a certain point in her demonstration Moss states that “your scoring rule and your actual credence in a proposition p determine the *expected epistemic value* of your having a particular credence in p ” (Moss 2012, 4), but regrettably, Moss doesn’t take this to imply that assigning a particular credence to p may already be an expression of these two factors – your scoring rule for p and the credence you have assigned to p - taken together.

⁶⁹ It doesn’t seem obvious that agents would literally be indifferent to small differences in credences; popular examples of scoring rules are generally strictly monotonic functions of credences scored.

The idea that credences alone do not express the actual epistemic value of propositions because they only refer to estimated *truth* value and not to *epistemic* value is, consequently, problematic. To relate to the conjunction of p's credence *and* the epistemic value of p's credence according to a scoring rule is to weigh what we typically assume has already been weighed in arriving at prior estimations of truth value that are expressed probabilistically in the credence assigned to p.

Furthermore, what I am suggesting here seems to also extend naturally from the conventional way according to which we understand credences and what they represent. There is a probabilistic range, $0 < C < 1$, from within which credences are assigned. As a matter of convention, what a credence measures is the likelihood that a select proposition is true. And the credence that is granted to p out of the probabilistic range represents the level of confidence in p being true. What *significance* a credence can have for an agent, independently of the agent's assessment of p's credence, does not seem relevant. At least not to an assessment of how confident one ought to be in believing p. If a proposition's level of credence is the indicator according to which one can assess how likely it is that p, then p's credence is what one ought to go by. And if one has two diverging credences to go by – for instance, the credence that A has assigned to p and the credence that B has assigned to p, then it is in reference to these credences that p's credence ought to be arrived at, if, that is, an estimation of the likelihood of p is what is being questioned. This seems to be the assessment that proper epistemic compromise wants to attain.

In sum, there seem to be two different matters to consider here. The first relates to how likely it is that a particular proposition is true, where this is commonly expressed by credences in the form of probabilities. The second matter relates to what it means, from

the point of view of preferences regarding diverse credences, for a particular proposition to have a particular credence. These seem to be two distinct questions, only the first of which seems relevant to an actual assessment of the designated proposition's likelihood.

EPISTEMIC COMPROMISE AND THE AGENTS TO WHICH IT APPLIES

Moss' discussion of epistemic compromise and her proposed solution focuses on agents that are somewhere between idealized agents and ordinary human believers. The sense in which these agents are idealized is that their scoring rules trade on a purely epistemic currency. Meaning that the way in which they value credences is a function of their (epistemic) preferences with regard to the implications of the proposition being true, as opposed to any other kind of non-epistemic preferences. The sense in which these agents are more human than idealized is the assumed variation, between agents, in the epistemic value attributed to credences. In this sense the fact that "different agents may value closer estimates differently" (Moss 2012, 3) means that even though their preferences relate to epistemic factors, they may nonetheless vary between agents.

But if these were ideal agents, conceived of as some kind of epistemic *homo economicus*, or as *Econs* (Thaler and Sunstein 2008), we would likely assume that they would not only condition on evidence in the same way, but that they would also be impartial to whether the difference between varying credences relates to credences that are closer to certainty or further away from it. Credence is, after all, a probabilistic expression of a particular measure of epistemic viability that is determined by where in the probabilistic range - $0 < C < 1$ - a debated proposition is located, in relation to a particular body of evidence. Moreover, what makes for the value of credence is the relation it bears to the actual truth

value of the proposition under consideration; and conventionally, the closer it is to certainty the greater value it has. And it seems plausible to assume that these idealized, non-human, agents would value credences *purely* epistemically, such that closer estimates would *always* be better because they would represent greater accuracy. And in this sense, epistemic value would, for such idealized agents, perhaps also be a superfluous notion. Once conditionalization on evidence is done in the same way for all agents, credences would also be valued in the same way, in accordance with their proximity to certainty. And in this sense such agents will share uniform scoring rules.

This observation about the kind of agents that Moss is referring to seems important. It is important because if scoring rules vary between agents, as do the epistemic values that these agents attribute to credences, then there must be grounds for such variation – there must, in this sense, be a reason why an agent values credence in one way rather than another. Furthermore, if an agent's preferences for certain credences over others are to be grounded in “purely epistemic concerns,” then the grounds for such variation must also be epistemic, and it isn't clear that this is something that Moss' account can accommodate.

EPISTEMIC VALUE CONSIDERATIONS

Is epistemic value *epistemically* relevant? Moss believes it is. She observes that different agents may value “closer estimates” differently; “you may value having 0.9 credence in a truth *much* more than having 0.8 credence, without valuing 0.6 much more than 0.5.” Moss takes such differences in epistemic evaluation to be *epistemically* significant. And it is this assumption that I wish to challenge.

By epistemic value what Moss has in mind is a preferential framework relating to the probabilistic evaluation of truths. Within this framework some probabilistic evaluations are preferred to others. Notably, Moss contends that this preferential framework is epistemic; suggesting that the type of preference for some probabilistic values over others is an *epistemic* preference. Setting aside for the moment *whether* different agents *do* in fact value credences differently, it seems important to ask whether *the fact* that they do so is *epistemically* relevant when considering the appropriate revisionary tactic for epistemic compromise. More specifically, the question is whether *the fact* that agents tend to value credences differently is relevant to an assessment of what *ought* to be the appropriate revisionary response regarding disputed issues between epistemic agents. This, after all, is what epistemic compromise is conventionally supposed to achieve.

In responding, I will start by addressing a practical problem. If different agents can value the same estimates of truth (credences) in different ways, then comprehending these truth estimations will depend on the epistemic values assigned to them by particular agents, set in place by varying scoring rules. And these scoring rules are what will be needed for deciphering the *actual* truth values of the propositions in question. If, for instance, 0.4 means Φ (some epistemic value) to me and Ω (some epistemic value) to you, our *actual* truth estimates will be $\Phi(0.4)$ and $\Omega(0.4)$, respectively. And this means that if we want to understand the credence assigned by each agent we need to abandon the accepted norm and notation of probabilistic evaluations. And seeing as we each grant the arithmetic expression assigned to P a different epistemic value, then epistemically, the same arithmetic expression means one thing for me and another thing for you. And this is problematic, since we no longer have any common way of assessing probability.

But aside from what the epistemic value model entails for our conventional norms for quantifying uncertainty, another matter we must consider in assessing whether epistemic value is relevant for epistemic compromise is whether the epistemic value on which this model is based can unproblematically be regarded as epistemic. I contend that it can't.

In ordinary real-world assessments under uncertainty, the significance that credences carry may vary in accordance with the kind of consequences that they represent. If, for instance, the truth of a proposition that has what we may regard as significant consequences is debated, a greater measure of caution might be taken in assessing its likelihood, because of the expected cost of being wrong. Alternatively, a more conservative estimation of the proposition's feasibility might be brought into play. And similarly, if the debatable proposition – for instance, a fact or an event – is marginally significant, less caution may be exercised in assessing its likelihood, because being wrong in such a case may not lead to any such substantial outcome. More generally, an agent might prefer credences that are closer to certainty on some occasions because their truth, in relation to a particular proposition, may be practically or emotionally significant⁷⁰. But similarly, on other occasions it is the possibility, rather than the probability of an event with grave consequences that will determine its significance, even if that event's probability is low.

The consequences of a possibility may often determine its epistemic value; the consequences of a particular proposition being true may lead to certain credences being treated otherwise than they would were they considered in relation to a different

⁷⁰ On the prevalence of this phenomenon see (Slovic et al. 2005; Slovic et al. 2007; Slovic 2010).

proposition being true. In the case of a possibility of grave consequences such as a nuclear attack, even low credences may be granted high epistemic value if the consequences are high.

And if, in such a case, we were seeking the consequential significance of particular credences in relation to a particular proposition, then averaging epistemic value to arrive at epistemic compromise might be the right way to go about it. If as part of the decision making process competing options were considered in relation to the pending threat, then compromising on the epistemic value of threat's probability may be feasible. In such a case what we would arrive at would be the significance of credence X for agent's A and B in relation to proposition P . But then it would appear that in this case we may have arrived at a compromise, but not an epistemic one.

On any plausible understanding of it, an *epistemic* compromise entails the incorporation of epistemically relevant considerations in accordance with a higher-level truth-eliciting revisionary strategy that doesn't jeopardize what epistemic compromise is all about. Ultimately, epistemic compromise aims to reach the most appropriate conclusion given the competing epistemic assessments, relating to how likely it is that a debated proposition is true, not to what it means for such a proposition to be true. In view of this, if we are going to profitably adopt the claims that Moss defends we should perhaps set aside scoring rules in favor of some more exact tool; in particular, we should adopt rules that assign different accuracy scores to credences in propositions depending on the subject matter of those propositions.

RATIONAL AESTHETIC DEFERENCE

Abstract: Is aesthetic deference ever rational? In this paper I argue that it can be. I defend aesthetic deference by means of a distinction between two kinds of aesthetic knowledge, predicative aesthetic knowledge and appreciative aesthetic knowledge. Whereas predicative aesthetic knowledge can be transmitted through deferential belief-formation, it is widely believed that appreciative aesthetic knowledge cannot. Building on this distinction and focusing on the relations between these two kinds of aesthetic knowledge I propose conditions under which aesthetic deference can be rational. Nonetheless, I suggest that there may be non-epistemic norms that undermine deferential aesthetic belief formation.

INTRODUCTION

Rational aesthetic deference becomes apparent when one person's aesthetic belief gives another person a *reason* to move his own aesthetic belief in the direction of the other person. It occurs when one person's aesthetic belief – for illustrative purposes let this be *my* belief – gives another person – for illustrative purposes let this be *you* – a normative reason to move your belief in the direction of mine, on epistemic grounds. In such a case what the first person believes also provides a justification for the second person's aesthetic belief. This kind of justification is an indirect justification, because it is based on reasons that merit deferring to someone else's judgment, rather than on reasons that support that judgment⁷¹.

⁷¹ In the remainder of the paper I use the notions of belief and judgment interchangeably.

There are plenty of examples from everyday life that show that aesthetic deference is something that we do. I might, for instance, go to the see a movie because a friend told me it was beautiful or buy an album because a colleague at work told me it was wonderful. Ordinary experience also suggests that in general we are also fairly comfortable with the idea of aesthetic experts whose aesthetic judgments are often considered prescriptive from a normative point of view. What a well known and esteemed art critic thinks about an exhibition, a book, or a dance piece, whether for instance it is fabulous, pretentious, striking, or kitsch, may impact our aesthetic beliefs about such things and determine what we do in relation to them – whether for instance we go to see the exhibition or read the book. There are a wide variety of additional familiar examples that one could think of along these lines. But these will suffice for claiming that deferential belief-formation is widespread.

Nonetheless, while it may be common for us to defer our beliefs on aesthetic matters, this does not mean that doing so is rational. In fact, there are reasons for thinking that aesthetic deference is not at all like ordinary kinds of epistemic deference which, under certain conditions, can be unproblematic. It is not, for instance, immediately obvious that deference on matters concerning the aesthetic value of a work of art, a building, or a piece of music is the same as deference on matters such as what time it is, the name of the capital city of Morocco, or any other such purely epistemic concerns in which knowledge acquisition is extended from one person to another person or source of information.

Most views on epistemic deference in the literature hold that it relates to the general conditions under which a person's beliefs can extend beyond the information that is directly accessible to that person, which principally relates to the extent that a person can

rely on a source of information which is not her⁷². But different authors take this to imply different things. Some take epistemic deference to involve knowledge acquisition processes in which a subject (the deferrer) relies on an external source's (the deferee) testimony in order to extend her knowledge to facts with which she has no direct acquaintance. Others take it to relate to cases in which "one person uses the deliverances of some information source, perhaps the opinions of another person, as a model for what to believe" (Joyce 2007, 187). Some have gone beyond epistemology and taken epistemic deference to imply the substitution by a decision maker of someone else's judgment for his own (Horwitz 2007). But it is generally true to say that epistemic deference relates to cases in which a person has a *reason* to move their belief in the direction of the belief of another person or information source when those are justifiably regarded as being in an epistemically advantageous position in relation to a particular matter at hand, either because of their capabilities in arriving at truths on such matters or because of the access they have to relevant information. A *reason* to defer to another person's judgment will typically be either that the person or information source in question has access to or understanding of some relevant knowledge or because of their judgmental capabilities in the relevant areas. The core question surrounding epistemic deference characteristically relates to what constitutes a reason for difference, which in turn addresses the conditions for epistemic reliability and knowledge extension.

⁷² There does not appear to be a single view regarding epistemic deference in the philosophical literature. Instead there seem to be a number of closely related discussions. See for instance: (List *unpublished*; Joyce 2007; Horwitz 2007; Goldberg 2009; Kelly 2005; Keren 2007; Owens 2000; Steup 2001; White 2005; Christensen 2004; Pillar 2001; Fricker and Cooper 1987).

Yet the problem with aesthetic deference is that its plausibility does not only seem to relate to the reliability conditions of extending knowledge, but also to the intrinsic possibility that aesthetic knowledge *can* be extended. The problem of aesthetic deference concerns deferential aesthetic belief formation and the significance of extending aesthetic judgment through a vicarious reliance on someone else's acquaintance with the object to which that judgment relates.

The problem with aesthetic deference and what I believe distinguishes it from epistemic deference, is the notion of acquaintance and the role that it is traditionally thought to have in belief formation on aesthetic matters. Aesthetic judgments are typically thought to require an acquaintance with their object. What this means is that they require a direct and unmediated experience of the object to which they relate. According to the acquaintance requirement, aesthetic judgments that relate to objects not on the basis of acquaintance but rather on the basis of testimony – namely aesthetic and non-aesthetic information conveyed indirectly by the declarative descriptions of others - are considered problematic⁷³. In the remainder of the paper I defend a position according to which aesthetic deference is nevertheless sometimes rational.

There are two important claims that are implied by this suggestion. The first is that there is such a thing as aesthetic reliability which can make a person's aesthetic judgment more

⁷³ It is worth point out that there is both a broad and narrow reading of the acquaintance requirement. The broad view suggests that acquaintance with the object is needed for aesthetic judgments to be valid. And in this case a person may be acquainted with the object without having an aesthetic acquaintance with it. So I may have seen the view but not considered it aesthetically. The narrow reading of acquaintance would require that for aesthetic judgments to be valid one must be aesthetically acquainted with their object.

or less reliable. The second is that aesthetic judgment need not be based on acquaintance. Nonetheless, I conclude by suggesting that there may be important non-epistemic considerations in light of which aesthetic deference is problematic, but that this has nothing to do with the nature of testimonial knowledge or the reliability of other people's judgments.

Addressing the first of these claims requires that we expound on what makes a person worthy of aesthetic reliability and trust, such that their judgments give us epistemic reasons to defer. For instance, is this person someone who possesses aesthetic knowledge to an equal or higher degree, or is it someone that is more capable of making aesthetic judgments? Addressing the second point requires that we consider whether we ought to defer to the judgment of such a person. It also requires that we consider to what extent this is possible if aesthetic judgments require acquaintance with their objects. In the next section I propose a characterization of the aesthetic deferee; that person that we might hold as deserving of trust and esteem with respect to their aesthetic judgments, and to which we may consequently consider deferring to on epistemic grounds.

AESTHETIC EXPERTS

If there are people whose aesthetic judgments merit deference, then they must possess some kind of epistemic advantage in the realm of aesthetics. This advantage may be tied to the possession of aesthetic knowledge or experience, or to a capacity for making

aesthetic judgments. In each of these cases the person must be justifiably thought to deserve trust and make possible relying on their judgments⁷⁴.

So what does possessing aesthetic knowledge, experience, and judgmental capabilities amount to? Insofar as it relates to reliability, *aesthetic knowledge* relates to aesthetic judgments that have been reliably formed and express true justified beliefs about the aesthetic quality of a particular object. If John has aesthetic knowledge relating to Ω , he holds a reliably formed judgment about Ω that predicates an aesthetic quality α on Ω , and John believes this judgment to be a true justified belief about Ω . Saying that John has aesthetic knowledge about Ω means that he holds aesthetic beliefs about Ω , that he takes to be true and justified. And depending on his reliability I may take this knowledge as a basis of what to believe.

⁷⁴ A person need not be generally *superior* in these capacities to warrant our deference, they may be *locally* better in relation to some particular proposition in aesthetics. Additionally, such a person needn't be better than us at all, they may just be *as good as* us. As long as we have reasons to rely on someone's judgment, their judgments can impact our own judgments – either because they happen to be in an epistemically advantageous position or because we exercise caution by counter balancing our own judgmental verdicts by incorporating the verdict of other reliable people. Hence for our purposes it suffices that they be reliable, for instance that they count as peers. A peer, as opposed to a superior or an expert, is not better, but equal and this in itself makes them as reliable as oneself. Nonetheless, whether or not the person in question is superior – for instance, if they are an aesthetic expert, when you are not – will have normative implications with regards to your justification in terms of relying on their judgments and the weight that their judgments are granted. As a rule, the more superior they are in the relevant capacities, the more justified one will be in trusting their judgment and giving it greater weight.

It needn't be the case that I rely on *any* kind of aesthetic knowledge that a reliable person has, but instead I may only rely on a certain type of aesthetic knowledge or only on aesthetic knowledge that relates to a particular area – sculpture for instance, or performing arts, and not, perhaps, aesthetic knowledge about painting. Hence for purposes of relying on aesthetic knowledge it may be local aesthetic knowledge relating to a particular domain of aesthetics, as well as global aesthetic knowledge relating to overall aesthetic knowledge.

But there must seemingly be something about another person's aesthetic knowledge that makes it legitimate to *defer* to them. One possibility is that they possess aesthetic knowledge that you don't. Thus if John knows that “the chair is majestic” and you don't – perhaps you only know that it is ‘dignified’ and ‘lofty,’ or perhaps you don't know anything at all about it - then John is advantageously positioned when it comes to aesthetic knowledge about the chair.

A person may also have *experience* with aesthetic judgments so that this experience positions them in an epistemically advantageous position in relation to aesthetic judgments. Many writers seem to believe that experience privileges. And in aesthetics, experience also seems to play an important role in establishing a person's privileged epistemic position. If a person has experience in making aesthetic judgments about, say, works of art of such-and-such a kind, then their aesthetic judgments in this area may be valued. One might for instance assign greater weight to the beliefs of a person that has a wide experience with aesthetic judgments, for instance with both good as well as bad art, because one might believe that this kind of experience gives a more rounded and mature perspective on objects considered from an aesthetic point of view. So too, if a person is

experienced in some area in aesthetics, then their aesthetic judgments in this area may be given greater authority because we assume that the person in question is ‘speaking from experience’ and that they know what they are talking about. In this sense the view expressed by an experienced person may override a contrary view held by a person with no experience, even if both have made well founded aesthetic judgments. Hence there is a sense in which we value aesthetic experience – such as a familiarity with works of art, and a repertoire of looking at things from an aesthetic point of view. And in cases where aesthetic experience is *justifiably* valued, deferential belief formation may be warranted on these grounds.

Lastly, a person’s capacity for *making aesthetic judgments* may lead us to regard them as deserving of trust and esteem, such that if we regard someone as having a good judgmental capability on aesthetic matters this may give us reason to rely on their judgments in forming our own. So what might a commendable capacity for making aesthetic judgments amount to?

Firstly, someone may simply be good at making aesthetic judgments. For instance they may know to look at the right non-aesthetic (but aesthetically relevant) qualities – such as colors and composition. And they may know how to interpret these qualities in novel and insightful ways. Secondly, they may have a good sense for aesthetics, which provides them with a greater sensitivity for identifying aesthetic characteristics in ordinary and artistic objects. Thirdly, they may have a good performance record of making good aesthetic judgments. And by good aesthetic judgments I mean judgments that have convinced others, met consensual aesthetic standards, or given other people insight to see things from an aesthetic point of view.

In sum, if there are people whose aesthetic judgments merit deference, then they must possess some kind of epistemic advantage in the realm of aesthetics. This advantage may be tied to the possession of aesthetic *knowledge* or *experience*, or to a capacity for making *aesthetic judgments*.

Having established that if there are people whose aesthetic judgments merit deference they must either possess aesthetic knowledge, experience, or judgmental capabilities that one has reason to rely on, in the next section I discuss the acquaintance requirement for aesthetic judgment to see whether we ought to defer to the aesthetic judgment of such persons. In doing so I also offer a distinction between two kinds of aesthetic knowledge corresponding to two ways in which aesthetic judgments can be understood, and consider their aesthetic significance and how they relate to each other and to the acquaintance requirement.

ACQUAINTANCE

The Acquaintance Principle (Wollheim 1980) is a much discussed principle in aesthetics. The original formulation of the principle appears in a passage in Richard Wollheim's *Art and its objects* (1980):

Realism acknowledges a well-entrenched principle in aesthetics, which may be called the Acquaintance Principle, and which insists that judgments of aesthetic value, unlike judgments of moral knowledge, must be based on first-hand experience of their objects and are not, except within very narrow limits, transmissible from one person to another (Wollheim 1980, p.233).

The Acquaintance Principle can be taken to make claims about the expressive nature of aesthetic judgments and about their proper genesis. The central appeal of the acquaintance requirement is tied to a widely held belief according to which there is an appreciative factor that underlies aesthetic judgments and involves first-hand aesthetic appreciation. It is this appreciative factor that is assumed to be lost when aesthetic judgments are made indirectly, by deference, rather than acquaintance.

In view of the acquaintance requirement, the underlying concern with aesthetic deference is that the aesthetic judgments which are deferred to do not carry appreciation with them, and hence if one's judgments are based on testimony about some object's aesthetic qualities rather than on an acquaintance with it, they do not incorporate the appreciative dimension which acquaintance affords, and which is constitutive of proper aesthetic judgment. The acquaintance requirement suggests that what acquaintance enables is aesthetic appreciation, which is what aesthetic judgments express. And hence judgments based on other judgments, specifically, on judgments made by other people and not on acquaintance, do not carry appreciation with them and are thus improper.

APPRECIATIVE AESTHETIC KNOWLEDGE

The idea of an appreciative aesthetic knowledge should be familiar. It suggests that there is a kind of aesthetic know-how of *what it is like* to appreciate aesthetic properties as they are realized in aesthetic objects. The appreciative aesthetic knowledge of Ω is what it is like to appreciate Ω aesthetically. Consequently an aesthetic judgment such as " Ω is beautiful" expresses an appreciation of beauty as it is experienced in Ω . And an acquaintance with Ω is a necessary condition for being able to make such a claim. If the

gracefulness of a line in a painting can only be appreciated through first-hand experience, not through any kind of description of it⁷⁵, appreciative aesthetic knowledge can only be arrived at by appreciation.⁷⁶ Such a position has been widely endorsed in the literature. Paisley Livingston, for instance, suggests that “even the most genial descriptions cannot enlighten us regarding the specific splendors of works of Schubert, Villon, Balthus, *et al.*, if we have no prior acquaintance with these works (or adequate surrogates thereof). ... [moreover] someone’s descriptions can inform one perfectly well about how the work is surprising, but only through a first-hand, description-free experience can one fully gauge the work’s surprise value” (Livingston 2003, 276 –277). Others, namely Sibley (Sibley 1965), Robson (Robson 2012) , Budd (Budd 2003), and Hopkins (Hopkins 2011), share this view.

PREDICATIVE AESTHETIC JUDGMENTS AND EVALUATIVE VERDICTS

But appreciative aesthetic knowledge is not the only kind of knowledge that aesthetic judgments express. Our aesthetic judgments often also convey aesthetic information. Take the judgment that “ Φ is beautiful.” In addition to its conveying an appreciation of what it is for Φ to be beautiful, it also conveys information about Φ , namely, that it is beautiful. In this case it is still an aesthetic judgment about Φ , since it conveys what you believe about Φ from an aesthetic point of view. The statement “ Φ is beautiful” can be

⁷⁵ The characteristic analogy illustrating this is that a fully determinate description of a graceful line in a picture does not entail its being graceful (F. N. Sibley 1974).

⁷⁶ Paisley Livingston makes this especially crisp by differentiating between knowing and gauging: “someone’s descriptions can inform one perfectly well about how the work is surprising, but only through a first, description-free experience can one fully gauge the work’s surprise value (Livingston 2003, 277).”

understood predicatively, articulating the existence of a predicative relation between beauty, an aesthetic property, and Φ , the object to which that property belongs. And there doesn't appear to be any kind of insurmountable barrier to knowledge of something's being beautiful being transmitted from one person to another. In supporting this claim Malcolm Budd observes that "judgments of aesthetic properties are as transmissible from one person to another as are other kinds of judgment" (Budd 2003, 392).

The idea of predicative aesthetic knowledge suggests that some aesthetic judgments can be based on this kind of predicative or declarative aesthetic knowledge. The guiding thought here is that in the absence of acquaintance, holding aesthetic beliefs is possible in the same sense that, to borrow an example from Sibley, someone who has not heard a joke can still be justified in believing that it is funny (F. Sibley 1965). I may believe that a joke is funny because I have been told so, not only because I have experienced its humor⁷⁷.

But if predicative aesthetic judgments are aesthetic in any interesting sense they must seemingly express more than just information about a property, which happens to be an aesthetic property, possessed by some object. For this relation to express an aesthetic judgment it must also be normative; it must establish that the object in question possesses a property that is good, or commendable, from an aesthetic point of view.

⁷⁷ Nonetheless, aesthetic variations of this same example seem odd. As Robson observes: "Consider, for example, how odd the following statement sounds: 'It's such a wonderful novel; insightful and moving, with the most beautiful and bewitching language, it's such a shame I've never read it' (Robson 2012, 4). The next section will suggest what may be different between these two kinds of statements and why the statement set forth by Robson may appear odd.

One way that a statement expressing a predicative relation can be normative is if evaluative aesthetic judgments can legitimately be derived from it. And assuming the reliability of the person expressing the judgment – a reliability that is based either on their knowledge, or their competence, or simply on an aesthetic sensibility that they share – it would be unreasonable not to accept evaluative aesthetic verdicts based on them. In this sense it seems plausible that evaluative verdicts may be legitimately derived from predicative aesthetic statements.

But this is not to say that predicative aesthetic judgments, even those made by the most aesthetically esteemed person, can enlighten us regarding the specifics of aesthetic appreciation of an object if we have no prior acquaintance with it (or with any kind of adequate surrogates). What I am suggesting is that aesthetic judgments made by others can legitimately enable evaluative aesthetic verdicts to be derived from them. And while those verdicts will be legitimate aesthetic judgments, they will not give us insight into the specifics of what it is like to appreciate the object related to from the particular aesthetic point of view. For this to happen, acquaintance is needed.

I still want to consider why it is that acquaintance seems to play such an important role in aesthetic judgments, to the extent that some have thought that you cannot form aesthetic beliefs without acquaintance. I also want to consider why aesthetic appreciation cannot be transmitted. Does for instance the acquaintance requirement entail that appreciative aesthetic knowledge can *never* be transmitted, or is it simply difficult to do so?

RETHINKING ACQUAINTANCE

The acquaintance requirement appears to have something to do with an experiential dimension of aesthetic properties that is given first-hand, presumably by perception, and cannot be properly appreciated unless one is acquainted with the object in this way. Arguably, aesthetic judgments relate to properties that require this dimension. And the properties in question call for a first-hand perceptual relation that is satisfied by acquaintance and gives rise to the required appreciative aesthetic knowledge.

But if appreciative aesthetic knowledge is required because aesthetic properties need perceptual acquaintance to be properly appreciated, then the same should be true of judgments that relate to other kinds of perceptual properties, such as colors. Yet presumably, no one thinks that you cannot form the belief that something is red by deferring to the judgment of another reliable and competent person that this is so, even though in such a case you do not know exactly what it is like for Φ to be red without seeing Φ yourself.

A possible answer here might be that you do not need to be perceptually acquainted with Φ to believe that it is red because you know enough about red things to know what Φ being red means; perhaps because you assume that it is similar to other red things. But cannot the same be said of aesthetic properties? Can you not also say of Φ that it is graceful because you know enough about graceful things to know what Φ being graceful means? If I know enough about grace, why can't I know that something is graceful by deference?

Perhaps knowing that Φ is graceful by deference is different to knowing that Φ is red by deference, because whether or not Φ is graceful is more complex and is perhaps also a matter of taste, whereas knowing whether Φ is red is a matter of basic perceptual competence. And basic perceptual competence is something we can most readily assume that other people have, whereas an equal appreciation of grace is not something that it is as easy to assume that other people have, or at least we can assume that there is greater variation in opinions about what it means for something to be graceful.

But if we assume sameness of sensibility between deferrer and deferee can we not also legitimately defer on aesthetic matters? To deny this seems to deny too much. After all, rational aesthetic deference presumably occurs when a person has a *reason* to defer to the aesthetic judgment of another person. And the assumption of sameness of sensibility to the deferee is a plausible example for such a reason. You won't defer to the opinion of someone whose taste you do not value, but rather to someone whose taste you do value, to the extent that you can rely on aesthetic judgments which that person makes.

Hence it isn't clear why a judgment about perceptual properties such as color can be based on testimony whereas a judgment about an aesthetic property such as grace cannot. If what is essential in aesthetic judgments is perceptual acquaintance with the objects to which those judgments relate, then assuming sameness of aesthetic sensibility, judging that something is graceful by deference seems to be a genuine possibility that weakens the claim that one cannot form an aesthetic judgment without acquaintance.

However, if there is an appreciative quality that belongs to aesthetic judgments and this quality is not a perceptual quality, but is nonetheless a quality that must be appreciated by

acquaintance, then perhaps aesthetic judgments are simply judgments that one must arrive at for oneself. Before I proceed to explore the latter possibility, one last option should be explored relating to the difference between judgments about color and judgments about aesthetic qualities.

Consider the distinction between information which is purely aesthetic – e.g., ‘the building is beautiful’ – and non-aesthetic information which is nonetheless aesthetically relevant – e.g., ‘the structure is *symmetrical*,’ ‘the structure is *warm*.’ That the structure is symmetrical and warm may be why we think that it is beautiful. The latter kind of judgments might supervene on the identification of the former kinds of properties. And while deference about non-aesthetic information such as judgments about symmetry may be unproblematic, deference about purely aesthetic kinds of information may nonetheless require acquaintance. If judgments about non-aesthetic information which is aesthetically relevant are like judgments about color, then this might explain the difference between judgments about color and judgments about beauty. It suggests that ‘ Φ is red’ is a judgment about a non-aesthetic property which is aesthetically relevant but does not require acquaintance, whereas ‘ Φ is beautiful’ is a judgment about an aesthetic quality that does require acquaintance.

AUTONOMY

Moral judgments are sometime thought to be autonomous. The autonomy of moral judgments suggests that they cannot be borrowed from someone else, since moral

judgments are a matter that the individual must arrive at for themselves⁷⁸. In what follows I outline the autonomy argument as it is thought to apply to moral judgments and import it into our discussion to see whether aesthetic judgments also require autonomy. If they do, then perhaps it is autonomy, rather than acquaintance, that makes aesthetic deference implausible, even if the reliability conditions exist for aesthetic deference on epistemic grounds. If this proves to be a genuine possibility, then the problem of aesthetic deference does not so much relate to whether the deferrer can be in a strong enough epistemic position to attain aesthetic knowledge, but rather that there may be an important non-epistemic norm which, as Hopkins suggests, makes it illicit to form aesthetic beliefs based on other people's judgments (Hopkins 2011, 140).

In discussing autonomy in relation to moral expertise and deference, Julia Driver states that autonomy is often considered crucial to moral judgment because the person who borrows a moral judgment, holds it without grasping the *reasons* for that judgment (Driver 2006, 622). Borrowed moral judgments – that is, moral judgments that have been rationally deferred to – are based on indirectly justifying reasons instead of directly justifying reasons⁷⁹. They are based on reasons that merit deference, such as that someone who justifiably counts as a moral expert, superior, or merely possesses some kind of relevant epistemic advantage holds them, and not on reasons that appreciate the

⁷⁸ Anscombe believes that moral judgments that aren't arrived at in this way lead to a "bastard sort of morality, marked by heteronomy." (Anscombe 1981).

⁷⁹ In referring to borrowed moral judgments as moral judgments that have been *rationaly* deferred to, I wish to stress that even if there are epistemic reasons that support the epistemic legitimacy of those judgments, i.e., on epistemic grounds, these will still be indirectly justifying reasons for the deferred judgments, rather than directly justifying reasons.

normative force of the moral judgment itself. The autonomy requirement implies that moral deference is problematic because the autonomy of judgments requires that there be personal and direct insight into the reasons used to justify the moral judgment that one holds. The justification for a moral judgment must consequently be direct, rather than indirect (Driver 2006, 623–624).

This position can be broken down into two distinct senses in which autonomy is required for moral judgment. The first is that a person that holds a moral judgment must do so for the *right reasons*, namely those which justifiably support that judgment and allow one to be compelled by its normative force. The second involves a person needing to *make up their own* mind when it comes to moral judgments. The latter seems to be an important, albeit not a strictly epistemic norm for aesthetic belief formation.

It seems that whether or not aesthetic judgments are different to moral judgments in terms of the autonomy requirement depends on whether grasping the reasons for an aesthetic judgment is necessary for appreciating its normative force. And it does not seem obvious that grasping reasons has the same role in justifying aesthetic judgment as it does with moral judgment. In fact, failing to grasp the reasons for an aesthetic judgment does not normally appear to affect the justification of holding it (Driver 2006, 623). As we saw above, part of the appeal of the acquaintance principle is that one needs to appreciate the object's aesthetic qualities for oneself. And if grasping reasons was necessary for aesthetic judgment, this could purportedly be done without acquaintance with the object of judgment, just like in cases of moral judgments. So perhaps appreciating the reasons for a judgment is not a requirement for aesthetic judgment after all.

If we believe that indirect reasons for aesthetic judgment are not sufficient because of some kind of autonomy argument, then it is either because of the importance of grasping the reasons for the judgment for oneself or because of the importance of some kind of non-epistemic acquaintance in aesthetic belief formation, something along the lines of needing to make up one's own mind on aesthetic judgments. In light of the fact that making up one's own mind does not refer to perceptual acquaintance of aesthetic properties or to direct appreciation of reasons, we are left with a somewhat cryptic requirement.

In sum, while aesthetic deference can be warranted on epistemic grounds – such as that someone else's judgment may be sufficiently reliable to provide indirect justification for deferring to that judgment, it may nonetheless remain unreasonable because of some kind of non-epistemic norm, such as needing to make up one's own mind. It doesn't seem to me that there is much more to say about the latter condition, other than that it is the opposite of what deference suggests, namely not making up one's own mind, but relying on someone else. And perhaps this is its strength.

Abstract: The acquaintance principle (AP) and the view it expresses have recently been tied to a debate surrounding the possibility of aesthetic testimony, which, plainly put, deals with the question whether aesthetic knowledge can be acquired through testimony—typically aesthetic and non-aesthetic descriptions communicated from person to person. In this context a number of suggestions have been put forward opting for a restricted acceptance of AP. This paper is an attempt to restrict AP even more.

1. INTRODUCTION

There is a widespread view according to which aesthetic propositions of the form “X is beautiful” or “X is elegant,” can only be coherently regarded as a person’s beliefs if they derive from that person’s first-hand experience of the objects to which the predicates ‘beautiful’ and ‘elegant’ apply. This view has been articulated as the Acquaintance Principle (Wollheim 1980). The Principle maintains that it is implausible for a person’s aesthetic beliefs to be based on someone else’s aesthetic experience. Accordingly, while I may tell you that X is beautiful, you cannot come to believe that X is beautiful simply

⁸⁰For a published version of this chapter in *The British Journal of Aesthetics* see: (Konigsberg 2012). See also previous references to the Acquaintance Principle in the previous chapter.

because I believe it is or because I tell you so⁸¹. You must, so says the Acquaintance Principle, experience this beauty for yourself⁸².

I assume the reader will agree that this view has significant intuitive appeal. Aside from its claim regarding proper aesthetic beliefs, it also seems to express a widely accepted position regarding the subjective genesis of proper aesthetic appreciation⁸³. Moreover, it also appears to support the consensual anti-realist position regarding the metaphysical dependence of aesthetic properties on an aesthetically appreciative subject.

The Acquaintance Principle (AP) and the view it expresses have recently been tied to a debate surrounding the possibility of aesthetic testimony, which, plainly put, deals with the question whether aesthetic knowledge can be acquired through testimony – typically aesthetic and non-aesthetic descriptions communicated from person to person. In this

⁸¹ This reflects a distinction I make elsewhere between testimony as a source of aesthetic knowledge and testimony as granting a reason for rational aesthetic deference.

⁸² Another closely related view contends that aesthetic experience is only possible if its genesis is in a first-hand aesthetic experience. According to this view, it is plausible for a person to say that X is beautiful if and only if he has experienced the beauty of X for himself. These are two closely related but distinct views. The first, roughly, relates to beliefs, the second to experiences. The two views are interrelated yet should be kept apart so as to distinguish claims made about each. My focus here will be on aesthetic beliefs but ultimately I will also discuss aesthetic experience.

⁸³ This can be understood in a number of ways. One way that will not be considered here but is worth mentioning nonetheless, is that expressed by Michael Tanner: “judgments of aesthetic, and in some cases moral, value must be based on first-hand experience of their objects not simply because one is in no position to assert the presence of the requisite properties without experience, but also because one is not capable of understanding the meaning of the terms which designate the properties without the experience” (Bermúdez and Gardner 2003, 33).

context a number of suggestions have been put forward opting for a restricted acceptance of AP. This paper is an additional attempt to restrict AP even more than previous suggestions.

The paper is composed of two parts. In the first part I present and discuss AP, its underlying conceptual foundations, and a number of responses to AP and the view it expresses in the recent literature. This discussion paves the way to Part II, where I argue, by way of demonstration, that AP is varyingly plausible in different domains.

Part I

2. THE ACQUAINTANCE PRINCIPLE

The Acquaintance Principle appears in a passage in Richard Wollheim's *Art and its objects* (1980):

Realism acknowledges a well-entrenched principle in aesthetics, which may be called the Acquaintance Principle, and which insists that judgments of aesthetic value, unlike judgments of moral knowledge, must be based on first-hand experience of their objects and are not, except within very narrow limits, transmissible from one person to another (Wollheim 1980, p.233).

This passage, and the principle that it expounds, can be understood as making two distinct claims that it will be useful to note and distinguish at the outset. The first of these claims relates to aesthetic beliefs, the second to aesthetic experience.

With regard to aesthetic beliefs, the claim made by AP is that without aesthetic experience aesthetic beliefs are improper. As I understand it, the claim can be articulated roughly as follows:

- Aesthetic beliefs involve aesthetic predicates which can only be properly predicated if the manifestation of the predicate they are expressing – e.g., beauty, elegance – has been experienced first-hand.

This means that unless *you* have had first-hand experience of the designated aesthetic object you cannot uphold aesthetic beliefs about it. This restriction specifically applies to proper *aesthetic* beliefs, for which first-hand experience appears to be a condition. I emphasize *aesthetic* beliefs because there doesn't appear to be anything problematic with beliefs in other areas, outside the domain of aesthetics, being proper in the absence of first-hand experience. There doesn't for instance seem to be anything wrong in believing that it is raining without experiencing - seeing, hearing, smelling, the rain, nor is there anything wrong in believing a structure is unstable because someone has told you so⁸⁴.

With regard to aesthetic experience, AP seems to be stating that:

- Without direct acquaintance, aesthetic experience isn't proper.

This latter claim is expressive of a widely held view that aesthetic experience requires a direct and unmediated familiarity with its object, and that it is this kind of familiarity that is purportedly what gives rise to proper aesthetic experience (Budd 2003; Meskin 2007;

⁸⁴ On this consider what Mark Owen Webb has noted about the epistemic importance of ordinary testimony: "I do not have to rely on my own cognitive resources; I may freely borrow from the resources of others. If I do not do so freely borrow, I shall be hopelessly imprisoned in an impoverished set of beliefs about only those things which I have myself experienced and can remember" (Webb 1993, 261).

Livingston 2003; Hopkins 2000; Laetz 2008), on which proper aesthetic beliefs are based.

Let us put this distinction between aesthetic beliefs and aesthetic experience crisply: AP, as we understand it, states that:

- 1) Proper aesthetic beliefs must be based on first-hand aesthetic experience;
- 2) Proper aesthetic experience must derive from first-hand acquaintance.

As I have noted, these are two separate issues. Ultimately my purpose in this paper is to make a point about aesthetic beliefs. All the same, to make the point I wish to make, I will need to discuss the nature of aesthetic experience from which proper aesthetic beliefs derive.

2.1 AESTHETIC BELIEFS

Responses to AP in the recent literature have expressed two distinct concerns with regard to aesthetic beliefs. The first of these is epistemic in nature and is concerned with whether, in the absence of first-hand experience, aesthetic beliefs can be sufficiently *grounded*. The second may be regarded as essentialistic, and is concerned with whether *aesthetic beliefs*, as opposed to other kinds of beliefs, are proper in the absence of first-hand experience⁸⁵. This latter question emphasizes the particular nature of aesthetic beliefs and is concerned with whether what it is that aesthetic beliefs typically express

⁸⁵ A similar distinction has been made by Meskin, between psychological and epistemological issues with regard to aesthetic testimony: “There are two distinct phenomena that must be explained. The first is psychological: we do not accept aesthetic testimony to the same extent that we accept other sorts of testimony. The second is epistemological: aesthetic testimony does not have the epistemic value possessed by other forms of testimony. What explains these phenomena?” (Meskin 2007, 68).

can be captured without first-hand experience. Thus in the former case the question is largely whether indirect experience – typically testimonial experience - can provide sufficient grounds for aesthetic belief. In the latter case the question is largely whether aesthetic beliefs that are not based on first-hand experience are proper.

My aim here is both epistemic and essentialistic. I aim to show that aesthetic beliefs can be based on testimony, and that basing aesthetic beliefs on testimony can provide sufficient ground and does not necessarily entail compromising aesthetic beliefs in any way. But if this was my sole claim, I would not be the first to make it. Recent proposals have suggested that certain kinds of aesthetic beliefs can be legitimately acquired without first-hand experience (Meskin 2007; Livingston 2003; Budd 2003). Moreover, it has also been suggested that a particular kind of aesthetic knowledge – roughly, declarative aesthetic knowledge relating to the existence of certain aesthetic properties in an object – can be transmitted from person to person (Budd 2003). In what follows I propose to go further in limiting AP and argue, by way of demonstration, that it is not only possible to transmit declarative aesthetic knowledge through testimony but that aesthetically appreciative experiences – to be more precise, a particular kind of appreciative aesthetic experience - can be transmitted by testimony. But before I proceed to argue for this claim let me show how my proposal is located in the context of the current views in the literature.

2.2 WEAKER VERSIONS OF THE AP

Recent discussions (Meskin 2007; Livingston 2003; Hopkins 2000; Budd 2003) have offered revised proposals of AP in which some of its tenets are maintained and others restricted. One positive amendment that these discussions have proposed suggests that in

the absence of first-hand experience and under certain epistemic restrictions aesthetic judgments *can* provide knowledge – declarative knowledge – which is the type of aesthetic knowledge that can be conveyed by descriptions pertaining to the presence of an aesthetic property in an object. In so doing these views have been defending the possibility of aesthetic beliefs, epistemically understood, in the absence of first-hand experience. This defense largely appeals to the possibility that declarative aesthetic beliefs can be sufficiently grounded even in the absence of first-hand experience.

The view, which at root I take to be shared across recent accounts, is that declarative knowledge can be epistemically grounded even when it is not based on first-hand experience, but rather on testimony, that is, on information transmitted from one person to another. The guiding thought here is that aesthetic beliefs in the absence of aesthetic acquaintance are possible in the same sense that, to borrow an example from Sibley, someone who has not heard a joke can still be justified in believing that it is funny (F. Sibley 1965). For the sake of convenience let us call the aesthetic knowledge that can be proper without first-hand experience *predicative knowledge*. And let us call the aesthetic knowledge that the literature concedes cannot be proper without first-hand experience *aesthetic appreciation*. I will elaborate further on this distinction in the following pages.

This differentiation between aesthetic appreciation and aesthetic predication purportedly makes aesthetic testimony, as a source of aesthetic knowledge, plausible in a limited way. This distinction between two types of aesthetic knowledge, one of which makes aesthetic beliefs in the absence of first-hand experience rationally permissible, has been expressed

in various ways⁸⁶. Malcolm Budd for instance claims that “*there is no insurmountable barrier to knowledge of something’s being beautiful being transmitted from one person to another*” (Budd 2003, 387). This can be taken to mean that Budd understands qualifications such as ‘X being beautiful’ predicatively, and thus as permissible. Elsewhere Budd states that “an item’s gracefulness, in contrast to its being graceful – likewise, an item’s beauty, unlike it’s being beautiful – cannot be transmitted from person to person through testimony. ... *Although aesthetic judgments do not carry appreciation with them, judgments of aesthetic properties are as transmissible from one person to another as are other kinds of judgment*” (Budd 2003, 392). Here Budd emphasizes the above-mentioned distinction. To use our terminology, he states that predicative aesthetic knowledge as opposed to appreciative aesthetic knowledge is transmittable via testimony. In a similar vein, Paisley Livingston states that “True descriptions can reliably inform us that a joke is funny and that an action film is brutal and stupid, and given sufficient sameness of sensibility, it would be unreasonable not to accept an evaluative verdict based on them; but even the most genial descriptions cannot enlighten us regarding the specific splendors of works of Schubert, Villon, Balthus, *et al.*, if we have no prior acquaintance with these works (or adequate surrogates thereof ... *someone’s descriptions can inform one perfectly well about how the work is surprising, but only through a first, description-free experience can one fully gauge the work’s surprise value*” (Livingston 2003, 276 –277). Hence Livingston endorses the same distinction and accepts the permissibility of predicative aesthetic knowledge.

⁸⁶ On a related matter compare Sibley on a similar distinction between aesthetic judgments and the attribution of an aesthetic property. See: (F. Sibley 1965, 137).

Thus these citations suggest that recent approaches to AP appear to endorse the distinction between two types of aesthetic knowledge in roughly the terms I have suggested above. To recap, predicative knowledge does not appear to be distinct from other kinds of propositional knowledge. But aesthetically appreciative knowledge is a kind of aesthetic know-how of *what it is like* to aesthetically appreciate the aesthetic properties as they are realized in the aesthetic object (Budd 2003). The recent literature suggests that aesthetic knowledge of the first kind is possible even without first-hand experience, that is, on the basis of testimony. While aesthetic knowledge of the second kind is not considered possible in the absence of first-hand experience. Thus notwithstanding the amendments that have made in the recent literature there is a consensus among these views that without first-hand experience one cannot fully gauge an object's aesthetic value, appreciatively conceived. The latter, it is maintained, requires a first-hand, description-free appreciative experience⁸⁷. In this vein it is generally accepted that the gracefulness of a line in a painting can only be appreciated through first-hand experience, not through any kind of description of it⁸⁸.

I relate to this last example because it exhibits a concern, expressed in the aesthetic literature, relating to the inability to fully appreciate the aesthetic merits of a perceptual object without perceiving it. I note this here because while the views referred to above have done well to weaken AP, the position they share still appears to accommodate a

⁸⁷ Paisley Livingston makes this especially crisp by differentiating between knowing and gauging: "someone's descriptions can inform one perfectly well about how the work is surprising, but only through a first, description-free experience can one fully gauge the work's surprise value (Livingston 2003, 277)."

⁸⁸ The typical analogy illustrating this is that a fully determinate description of a graceful line in a picture does not entail its being graceful (F. N. Sibley 1974).

particular ambiguity in the conception of first-hand experience – namely, the aesthetic experience they are referring to seems, by and large, to be *perceptual experience*.

The relation between aesthetics and perception is a broad topic dealing with historical, conceptual, sociological, cognitive, and psychological questions⁸⁹. It is not my aim to address any of these here. Instead I wish to suggest that the weakening of AP by recent accounts has still kept first-hand *perceptual experience* with an aesthetic object as a requirement for an aesthetic belief to be proper. Yet I believe there are aesthetic experiences where perception is not a necessary component and consequently neither is it necessary for aesthetic beliefs, and possibly even knowledge. Moreover, I think that this is true both of predicative knowledge and of appreciate knowledge. That is, I think predicative aesthetic beliefs can be proper in the absence of perception, as can appreciative aesthetic beliefs. The first claim is straightforward, the second less so. Thus I think that the notion of aesthetic experience in the literature is still rather under-explained. But I contend that with the appropriate demonstration the conception of first-hand aesthetic experience can be broadened to include not only perceptual experience. It seems to me that a person *can* have first-hand experience of an aesthetic object without perceiving this object, and I think that we often take this for granted in our aesthetic conduct.

⁸⁹ Relevant literature on the topic includes: (Binkley 1977; Shelley 2003; Carroll 1999; Shelley 2003; Margolis 1960; Carroll 2004; Hopkins 2007).

2.3 AESTHETIC AUTONOMY

The AP exhibits an entrenched conception which stems from a multifarious tradition of subjectivist aesthetics that spreads over a plethora of works, traditions and approaches. I believe AP expresses an underlying conception of what I shall refer to as *aesthetic autonomy*, a stipulation according to which the aesthetic agent is someone who arrives at his aesthetic beliefs on his own, through his first-hand experience of the realization of aesthetic properties in an object⁹⁰. The said view also concedes, even if this is not always declared explicitly, that the agent acquires his aesthetic beliefs by *perceiving* the aesthetic properties in the object of his aesthetic appreciation. Moreover, in the absence of any claims to the contrary in the recent literature I will assume that this is the only plausible way in which it is considered possible for an agent to have aesthetic beliefs.

I do not want to completely depart from this conception of aesthetic autonomy nor do I want to deny the AP. All the same I would like to suggest a modified understanding of aesthetic autonomy, which will also bear on the AP. According to what I shall suggest, an aesthetically autonomous agent is required to have first-hand experience of the aesthetic properties that his belief predicates, and this is also a condition for his belief to be proper. In this I will be consistent with the widespread understanding of the AP. But I shall also want to claim that such an agent can experience aesthetic properties through testimony. That is, I will be arguing that a person *can* have an aesthetic experience – in fact, he can

⁹⁰ My use of aesthetic autonomy is different to the use that has been made of it by e.g., (Diffey 1982; Stecker 1984), where Aesthetic Autonomy is taken as an art for art's sake type of conception as oppose to an instrumental approach to aesthetic objects. My use of aesthetic autonomy makes reference to the autonomy of the individual in the constitution of his own aesthetic experience.

have an aesthetically appreciative experience - without having any direct perceptual encounter with the object to which the aesthetic properties that constitute that experience belong. If this suggestion turns out to be plausible this would mean that a person can be aesthetically autonomous without having to perceptually experience the object of his aesthetic appreciation.

2.4 AESTHETIC EXPERIENCE AND PERCEPTION

Seeing how prominent abstract and conceptual art is in our day, one would expect our conception of aesthetic experience to reflect what has become the norm. But at root, aesthetic experience still seems to be very much understood in relation to sense perception, as it was in the 18th century when the term was coined by Baumgarten⁹¹. And while nowadays it is common to appreciate the aesthetic merit of works that are not only perceptual in character, perceptual acquaintance – that is, first-hand perceptual experience - still seems to be considered a basic requirement of proper aesthetic experience. This position is quite clearly expressed by prominent defenders of aesthetic autonomy and the implicit connection made, for instance in the AP, between aesthetic acquaintance and perception. Kant is an early example. In *The Critique of Judgment* he

⁹¹ See: (Binkley 1977, sec. IV). In the debate surrounding testimony and the Acquaintance Principle, Meskin is an exception to what seems to be the prevalent conception linking the aesthetic experience with perception. Meskin notes that “it is simply not the case that aesthetic judgments are necessarily linked to *perceptual* experience” (Meskin 2007, 74). But he appears to be referring here to the aesthetic experience of non-artistic objects, those expressed in the aesthetic judgments made by scientists, mathematicians and philosophers. On the aesthetic appraisal of scientific discourse and discoveries see: (Kivy 1991); referenced by Meskin: (Meskin 2007, 74).

states: “Whether a dress, a house, or a flower is beautiful is a matter about which one declines to allow one’s judgment to be swayed by any reasons or principles. One wants to *see* the object with one’s own eyes, *as though one’s pleasure depended on sensation*” (Kant, sec. 8). Other, later, expressions include Alan Tormey: “In art, unlike the law, we do not admit judgments in the absence of direct or immediate experience of the object of the judgment. *We require critical judgments to be rooted in ‘eye-witness’ encounters*, and the indirect avenues of evidence, inference and authority that are permissible elsewhere are anathema here” (Tormey 1973). Frank Sibley, in his seminal *Aesthetic and non-aesthetic*: “It is of importance to note first that broadly speaking aesthetics deals with a kind of *perception*. People have to *see* the grace or unity of a work, *hear* the plaintiveness or frenzy in the music, *notice* the gaudiness of a color scheme, *feel* the power of a novel, its mood, or its uncertainty of tone. They may be struck by these qualities at once, or they may come to *perceive* them only after critics. *But unless they do perceive them for themselves, aesthetic enjoyment, appreciation, and judgment are beyond them*. Merely to learn from others, on good authority, that the music is serene, the play moving, or the picture unbalanced is of little aesthetic value; *the crucial thing is to see, hear, or feel*. To suppose indeed that one can make aesthetic judgments without aesthetic perception ... is to misunderstand aesthetic judgment (F. Sibley 1965, 137).” Last but not least in this list of prominent philosophers is Philip Pettit who takes aesthetic characterizations to be “essentially perceptual”: “the putatively cognitive state one is in when, perceiving a work of art, one sincerely assents to a given aesthetic characterization, is not a state to which one can have non-perceptual access” (Schaper 1987, 25). Pettit goes on to state that: “Aesthetic characterizations are essentially perceptual in the sense that perception is the

only title to the sort of knowledge – let us say, to the full knowledge – of the truths which they express” (Schaper 1987, 24–25).

I take these extracts from the literature to reflect a prominent and widely held conception of aesthetic experience as largely perceptual experience. Moreover, in the context of the view expressed in AP, I take this to mean that first-hand perceptual acquaintance is a necessary condition for aesthetic beliefs that are based on this experience, to be proper. Yet I contend that while most familiar aesthetic experiences are indeed perceptual, there also appears to be a sub-set of aesthetic experiences that do not depend on perception. My defense of the possibility of non-perceptual aesthetic appreciation stems from two empirical observations which I will assume that the reader shares. The first is that artistic objects are not the only objects that can be appreciated aesthetically. The second is that we often go through aesthetically conversational experiences that are instigated by other people – for instance friends, acquaintances, writers, critics.

2.5 THE SCOPE OF THE AESTHETIC DOMAIN

Before I continue a short clarification is needed. In what follows I wish to focus on aesthetics understood in the broad sense of the term, where experience rather than art is the axis of consideration. Aesthetics is ordinarily thought to be about art, or thought of as the philosophy of art. But aesthetics as I shall relate to it here is not exclusively about the artistic domain. In fact, the artistic domain seems to me to be part of the aesthetic domain, not exhaustive of it. To be more precise, I think aesthetics is actually about a particular type of human experience, the aesthetic experience. And I share the belief that the aesthetic experience “is largely elicited by artworks” (Binkley 1977). But I also believe that it is not only elicited by artworks. In fact, as I think the reader will agree,

aesthetic experiences of non-artistic artifacts such as aesthetic experiences of nature, buildings, people, mundane objects, theorems, and food are quite familiar in all wakes of life.⁹²

In part II of the paper I will make reference to examples taken from the Artworld. Yet my argument about the possibility of gaining aesthetic knowledge through testimony will apply to aesthetics conceived of broadly, which applies to non-artistic artifacts as well as to works of art. I will show how aesthetic autonomy is compatible with aesthetic testimony. I will demonstrate how a first-hand aesthetic experience is possible even when that experience is not perceptual. I will discuss several works of art whose aesthetic appreciation makes perceiving them inconsequential, and I will take the standard aesthetic appreciation of these works to demonstrate my claim. By attending to what an aesthetic experience of these works amounts to I shall pave the way to showing how their aesthetic properties can be aesthetically appreciated even when this experience is elicited through aesthetic and non-aesthetic descriptions communicated by testimony and not through sense perception⁹³. In so doing I hope to show that aesthetic experience is not equal to nor is it exhausted by perceptual experience.

⁹² Or at the very least language attests that they are. See also: (Zemach 1997; Kivy 1991) on other, non-aesthetic, forms of aesthetic appreciation.

⁹³ For more on the relation between non-aesthetic and aesthetic descriptions see: (F. Sibley 1965).

Part II

3. DOMAIN SENSITIVITY

I shall begin this section by making a claim, the truth of which I will demonstrate in the following pages:

- AP is domain sensitive. More specifically, AP is varyingly plausible in different domains⁹⁴. This means that: 1) in some domains it is not only *predicative aesthetic* knowledge that is possible without first-hand experience, but also *appreciative aesthetic* knowledge. 2) In some domains the perception of the aesthetic object appears to be inconsequential to its aesthetic appreciation.

Let us now attend to some examples from the domain of modern art, where I believe a sub-domain of aesthetic works supporting the above claim may be found.

3.1 ERASING THE NEED TO SEE

Robert Rauschenberg was fascinated by Willem De Kooning, and in 1953 he asked the artist if he could erase one of his drawings as an act of art. The genesis of Rauschenberg's

⁹⁴ Meskin has made a similar claim to this: "I will also argue that the epistemic status of aesthetic testimony depends on the aesthetic domain that is being considered. Consider the following sorts of things about which aesthetic judgments are made: paintings and sculptures, landscapes, faces, proofs, theories, souls, and sounds. The status of testimony is not the same in all of these domains. For example, the situation with respect to aesthetic testimony about works of art is rather different from the situation with respect to testimony about nature as well as certain abstract objects such as proofs and theories" (Meskin 2007, 69).

project is well documented: Rauschenberg went over to De Kooning's studio and said he'd like to erase one of his drawings as an act of art (Katz 2006). De Kooning, apparently intrigued, selected a multimedia work on paper that he knew it would be quite difficult to eradicate. And he was right. It apparently took Rauschenberg one month to get the sheet relatively clear of marks.

In *Erased De Kooning Drawing* (1953), Rauschenberg presents an almost white canvas with faint traces of ink and crayon. The canvas denotes the absence of a work of art that was previously there. It is this alluding to what is absent from the canvas that is in a large part what makes *Erased De Kooning Drawing* (1953) a work of art. It is a conceptual work of art. We know that it is a conceptual work of art because you have to know that there was a real De Kooning there that was erased. Moreover, the fact that the work that was erased belonged to one of the most regarded artists of the time contributes to the aesthetic impact of Rauschenberg's piece.

To seal his artistic performance, Rauschenberg asked artist Jasper Johns to inscribe the lettering in a box at the bottom of the picture. The box states: "Erased de Kooning drawing, Robert Rauschenberg, 1953." Now, let me be blunt: in point of fact, what we have here are three renowned artists collaborating on a work which is what we might - perhaps anachronistically - call a non-work. And by the fact of their artistic collaboration this non-work becomes a work of art.

3.2 NON-WORKS OF ART IN THE ARTWORLD

For many years now it has been widely recognized that in the contemporary Artworld esteemed curators, collectors, and artists have a special Midas-like ability to call attention to a work and by this very act inaugurate it as belonging to the Artworld and

consequently endow it with aesthetic value. At other times it is not *who* the person is that does the artistic inauguration but rather *the context* within which the work is located that establishes it as a work of art (Weitz 1956; Danto 1964). For instance, nowadays it doesn't seem at all farfetched to think that if one were to place a spoon, or an old rug, a broken iphone, or a half-full coffee mug, or a car tire, or a hose, or a bed, or a desk in an empty room in a gallery, one's doing so would make viewers artistically intrigued. A common response by the observer in such scenarios might be to inquire into the background knowledge of the creation of the work as a means of grasping the non-perceptual meaning, or the significance, of the perceptual object.⁹⁵

Assuming that we accept that this is how things are in the Artworld, or that we at least accept that this is how they are in the Artworld's present state of evolution, then it seems that we have a choice. Either we accept that things need not be aesthetic to be works of art, or else we contend that aesthetics is a notion that denotes more than only what can be perceived. If the reader feels this choice is unfair, or instead feels a resistance to accept either of these options, this is understandable. Yet whatever the reader's artistic persuasion happens to be, to deny that works such as Rauschenberg's and the flourishing tradition of conceptual art that emerged in the second part of the 20th century is art, is to deny too much⁹⁶. It involves denying the dynamics of artistic production, consumption, and economic exchange in the Artworld during this period.

⁹⁵ Following Wollheim we might say that this background knowledge is the 'cognitive stock' necessary for appreciation (Wollheim 1980, 185–204), cited in (Livingston 2003, 267).

⁹⁶ Noel Carrol shares this belief, although he refers to non-perceptual art as non-aesthetic art: "It seems that the existence of anti-aesthetic art is a fact of the artworld and has been for some time ... The aesthetic theorist cannot stipulate what she will count as facts in the face of massive amounts of countervailing

But it is not my aim here to convince the reader to embrace a liberal conception of art or aesthetics. Instead, let me actually point out that to agree that the realm of aesthetics goes beyond what can be appreciated through perception is not to contend that all art is good art or that non-perceptual art ought to be regarded positively. To concede that more than perceptual appreciation is involved in the aesthetic evaluation of some of today's prominent works of art is simply to recognize that experiencing art has become more than merely a perceptual experience. I share James Shelly's contention that it is nowadays quite possible to have aesthetic experiences of imperceptabilia (Shelley 2003). Yet my aim here is not so much to argue that this is possible but more to demonstrate that there are aesthetic experiences in which perception is unimportant.

3.3 BACK TO THE DRAWING BOARD

To lay bare the aesthetic unimportance of perception in the aesthetic appreciation of some works of art let us return to *Erased De Kooning Drawing*. I think the reader will agree that no important information about Rauschenberg's piece is presented in the way it looks, except the fact that looking at it is artistically unimportant. To see this, consider how bizarre it would be to come up close and appreciate the lines and smudges created by the previous inhabitants of the canvas and now erased ink and crayon that make up *Erased De Kooning Drawing*. It would simply be a mistake, an embarrassing one come to think of it, to search for aesthetically interesting visual marks, lines, patterns, or shapes in *Erased De Kooning Drawing*. This is not to say that such aesthetically interesting evidence, which continues to grow daily" (Carroll 1999, 182). See also Shelly's discussion of rejections, such as Beardsley (Lamarque and Olsen 2004, chap. 6), of readymades and other non-perceptual artworks (Shelley 2003, 367).

patterns and shapes could not have haphazardly formed on the canvas by the erasing of the previous media. Nor does it mean that if they were observed such patterns could not be given symbolic meaning. It is merely to say that if one were to spend time considering the patterns left over from De Kooning's painting instead of appreciating the painting's absence, one would, aside from doing something non-standard and strange, be missing the point. The point here is that to appreciate *Erased De Kooning* as an artwork is not to appreciate its perceptual qualities⁹⁷.

To put things bluntly, *Erased De Kooning Drawing* is a relic of the traditional, perceptual, conception of artistic meaning (Binkley 1977, 81). And if there is aesthetic value to De Kooning's piece, it resides in the background knowledge we have about it, not in what we *see* in it. Such background knowledge can for instance be the knowledge that there was previously a painting where there is now an almost blank canvas; that the painting was a De Kooning; that it is now erased; that Rauschenberg erased it; and that even though Rauschenberg erased it and it is no longer there visually, it is still with mount and frame and positioned in a prominent place on a gallery wall; a wall which one must pay to view. Now in light of all of this let us ask: does one really *need* to see *Erased De Kooning Drawing* to appreciate its aesthetic merits? In considering this question, think of the information that you, the reader, now have about *Erased De Kooning Drawing*. Does this information – conveyed to you through this author's words – elicit an aesthetic experience of the piece? Could it? If it does give rise to an aesthetic experience,

⁹⁷ Compare Dickie's discussion of *Fountain*'s perceptual qualities: "Why cannot the ordinary qualities of *Fountain* – its gleaming white surface, the depth revealed when it reflects images of surrounding objects, its pleasing oval shape – be appreciated? It has qualities similar to those of works of Brancusi and Moore which many do not balk at saying they appreciate" (Dickie 1974, 42).

is this the same kind of experience you would have through a perceptual experience of the work? If it isn't the same kind of experience, is the difference between the two *essential* to the aesthetic appreciation of *Erased De Kooning Drawing*?

3.4 FOUNTAIN: THE AESTHETIC UNIMPORTANCE OF PERCEPTION

Let us keep these questions in mind while we look at another example, this time of a work that has gained its fame for not appealing to the viewer's perceptual appreciation of the perceptual properties that comprise it, Marcel Duchamp's *Fountain* (1917).

One of the notable differences between *Fountain* and *Erased De Kooning Drawing* is that the supposed object of *Fountain* – a urinal, was never designed or manufactured as an object for aesthetic appreciation. In this sense it is different from Rauschenberg's drawing. Indeed, in the case of *Fountain*, it is solely the fact that something which is clearly not a work of art has been placed in a context for aesthetic appreciation (Hanfling 1992) by a renowned artist that makes it into a work of art. It is effectively this act that transforms the urinal into something deserving of aesthetic appreciation.

Now assuming that we agree that *Fountain* is an object of aesthetic appreciation, we can now consider whether its aesthetic appreciation can be communicated via testimony, instead of needing to be experienced perceptually by the aesthetically appreciating agent. I tend to think it can. I believe this is so because nothing aesthetically important about *Fountain* seems to depend on one's being able to *see* it. All the same I recognize that this may be because I am considering *Fountain*'s aesthetic appreciation from what might be regarded as an essentialist point of view; that is, I am trying to excavate what is aesthetically essential in *Fountain* and omit any subsidiary experiential factors that may arise from a direct first-hand perceptual encounter with it. Let us keep this concern in

mind while we consider another two examples. We will readdress this concern in the last section of the paper.

Beforehand, and for the sake of keeping us on track, let me repeat what I take to be my core thesis: I think that it is possible to aesthetically appreciate something without perceptually experiencing it. I think that this is made possible because aesthetic experience can extend beyond the perceptual domain and can be arrived at in absence of perceptual experience by aesthetic and non-aesthetic information that is conveyed through testimony. Let us now consider two more examples of works of art whose aesthetic appreciation exhibits this thesis.

3.5 THE BASE OF THE WORLD AND VERTICAL EARTH KILOMETER

In 1962, Italian artist Piero Manzoni created *Socle du Monde* (*The Base of the World*), in which he exhibited the entire planet earth as his artwork. Manzoni placed a large metal plinth inscribed with the words *Socle du Monde* upside down in an open field in Herning, Denmark. The piece, positioned as a pedestal, announces that the world is a work of art.

Many have taken Manzoni's work to be conveying a message about the aesthetic nature of the world conceived as a work of art, a conception that seemingly renders the artist obsolete. Literally speaking one cannot view Manzoni's piece as a *whole* piece (if that is, one seeks to visually comply with the spatial extent of the work) unless one does so from The Moon or from some other location in space from which one can view the earth. All the same, the constraint of not being able to view the work as a whole does not appear to limit one's ability to aesthetically appreciate it.

Fifteen years after Manzoni created *Socle du Monde*, in 1977, a two-inch (five centimeter) thick solid brass rod extending one kilometer straight into the earth is

installed by American artist Walter de Maria in the Freidrichplatz Park in Kassel, Germany. The work is De Maria's contribution to the *Documenta* exhibition of that year. De Maria calls this work *Vertical Earth Kilometer* (1977). The full length of the rod is completely sunk into the ground so that only its very top which is level with the surface is visible. All that the viewer can see in *Vertical Earth Kilometer* is a sandstone square surrounding the brass rod's flat circular top which commemorates the depth of the rod below. The boring of the shaft in Kassel took seventy-nine days and went through six geographical layers. It is a continuous rod that is made up of combined lengths of brass each measuring 167cm.

The enormity of the kilometer of brass that has been sunk into the ground in Kassel exists as background knowledge in the mind of the viewer of *Vertical Earth Kilometer*. Because there is no visual trace of the kilometer of brass other than its exposed top, some have said that the work addresses the question of trust between artist and audience. On this interpretation the viewer sees a round disc which she takes to be the very top of the rod that is one kilometer long that has been plunged into the earth. The viewer assumes that what she sees is the top of the rod and it is *this* knowledge that supposedly elicits her aesthetic experience.

Sense perception does not seem to be essential for the aesthetic appreciation of the four works of art described above. This at least is what I would like to suggest. I make this suggestion in view of the fact that the perceptual dimension of each of these works does not seem to form a central part of the aesthetic merit of each of these works. Indeed, where perception does play a part in aesthetic appreciation, it does not seem to be of primary importance: Raschenberg's piece denotes the fact that it is an erased De

Kooning, and our knowing this seems to form the basis of that piece's aesthetic appreciation. Otherwise put, we do not typically appreciate *Erased De Kooning* because of how it looks. Similarly, Duchamp's *Fountain* makes the perceptual appreciation of the urinal inconsequential and invites the viewer to contemplate the artistic significance of the sanitary object in the non-standard context in which it has been placed⁹⁸. Moreover, *The Base of the World* is what we might call an aesthetic gesture; it does not so much stimulate visual perception but instead it suggests that the world, conceived of as a sphere that sits on a large metal man-made plinth, is a work of art. Finally, Walter de Maria's *Vertical Earth Kilometer* motivates a mental representation of something that cannot be perceived by observing what is visible in that work.

3.6 AESTHETIC AUTONOMY AND AESTHETIC TESTIMONY

I hope that these examples have convinced the reader that there is some art the aesthetic appreciation of which does not involve perceptual appreciation but rather another kind of, possibly conceptual, appreciation. Now the question that we appear to presently face is whether the fact that there are some works of art whose perceptual appreciation is not essential to their aesthetic appreciation should be taken to mean that perceptual appreciation is not necessary for the aesthetic appreciation of these works of art. Otherwise put, we need to establish whether the fact that seeing *Fountain* doesn't appear to be essential to grasping what lies at the heart of its aesthetic appreciation makes disregarding the perceptual dimension of *Fountain* aesthetically proper.

⁹⁸ Compare: (Dickie 1974, 42) & see footnote #7 above.

At first sight, the suggestion that I have made here seems to invite an objection. The objection is that the non-perceptual appreciation of any work of art that has a visual quality to it may be qualitatively different to the perceptual appreciation of that work. On this view, standing in the Freidrichplatz Park in Kassel looking down at the two-inch brass circle on the ground provides a different experience than that which can be elicited by the description of *Vertical Earth Kilometer*. Just as being told of what Rauschenberg did to create *Erased De Kooning Drawing* may elicit a different experience than seeing the piece would. This objection seems to appeal to common sense and may seem intuitively plausible. Its basic claim is that an experience of *Fountain* cannot be the same as a description of the experience of *Fountain* which would be conveyed by testimony, just as the taste of a fine steak cannot be transmitted from one person to another by a description of it. Let me try and get at what I take to be the basis of this claim.

3.7 CAN YOU APPRECIATE A STEAK WITHOUT EATING IT?

There appear to be dimensions of the aesthetic experience of a perceptual object that cannot be experienced without perception. For a start, even if every noticeable element of a perceptual experience is described and communicated, there are some elements that are formative to the experience but are not noticeable and hence cannot be communicated. If I tell you that a urinal of the type that is usually found in public lavatories is set on a plinth in the center of a dimly lit space in a gallery, I am not telling you enough, so the argument supposes, to elicit in you the kind of experience that you would have if you would see the urinal first-hand. Additionally, even if all communicable elements that belong to a perceptual object that is a work of art are communicated, they may not necessarily elicit the aesthetic experience from which they derive. This is because while

there may be non-aesthetic properties the specific arrangement of which can give rise (I dare say *cause*) to an aesthetic experience, this does not mean that if these same properties are arranged in the same way elsewhere on the basis of testimony, they will also lead to the same experience. Moreover, non-aesthetic properties – such as lines, composition, colors - can elicit different aesthetic experiences in different people and thus there is no reason to think that a description of the non-aesthetic properties by one person will elicit an aesthetic experience in another. Generally speaking, aesthetic experience seems often to be explained in reference to certain non-aesthetic elements, but the presence of these same elements does not necessarily yield the aesthetic experience which they explain. And hence even if you receive the most complete description of the aesthetic and non-aesthetic qualities of a work of art this may not elicit the same aesthetic experience or it may not in fact elicit any aesthetic experience at all⁹⁹. Moreover, one might even believe that the most perfect description could not convey the aesthetic experience of a perceived aesthetic object.

The heart of the objection is that there *is* a difference between perceptual and non-perceptual experiences. But this objection does not seem to me to be enough to deny that non-perceptual aesthetic experience is possible. What the objection suggests is that perceptual and non-perceptual experiences are different, and that the non-perceptual experience might possibly lack experiential dimensions that the perceptual experience has. In this sense the difference between both kinds of experience is understood as a difference in measure - the non-perceptual experience having less than its perceptual counterpart. Yet this says nothing about why the non-perceptual experience cannot be an

⁹⁹ See also: (F. Sibley 1965; F. Sibley 1959).

aesthetic experience. And thus in absence of such an argument I will assume the plausibility of the thesis I propose: non-perceptual aesthetic experience *can* be proper aesthetically appreciative experience. Moreover, non-perceptual aesthetic experience can ground proper aesthetic beliefs that convey appreciative evaluations of aesthetic objects.

4. CONCLUDING REMARKS

If X is a work of conceptual art, to claim that the aesthetic appreciation of X can be communicated via testimony is to say that the first-hand perceptual experience of X is not necessary for the aesthetic appreciation of X. Positively, it suggests that one can appreciate the aesthetic properties in X even if the source of one's experience is testimony rather than first-hand perceptual appreciation.

There seem to be two different positions that can be endorsed here. The first is strong, and suggests that a complete aesthetic appreciation of X (whatever that may be) can be attained through testimony and does not require first-hand perceptual appreciation at all. What this strong position is actually stating is that all the qualities that comprise the aesthetic appreciation of X can be experienced even when information about X is attained through testimony. This position takes all perceptual experience of X to be inconsequential to the aesthetic appreciation of it. And this seems to be far too strong a requirement and thus seems implausible.

The second position, which I endorse, is more moderate and suggests that a first-hand perceptual encounter with X can provide an aesthetic appreciation of X that cannot be transmitted through testimony. This suggests that factors such as the perceptual impact of a work of art and the actual feeling that one has by perceptually engaging with it are

exclusive to perceptual encounters. From this more moderate position we can agree that aesthetic experience based on testimony is not a complete aesthetic experience because it lacks a perceptual dimension. All the same we also contend that whether or not the perceptual dimension is crucial to the aesthetic experience of a work of art will depend on what is aesthetically essential in that work, which consequently depends on the nature of the aesthetic domain to which the work belongs.

Let me make a final note about convention. Even if we permit that non-perceptual aesthetic appreciation is possible, and consequently that beliefs grounded in such experience are aesthetically and epistemically proper, convention may still make the expression and attribution of aesthetic belief in the absence of first-hand acquaintance improper in the sense that a non-perceptual aesthetic belief may imply that a conventional first-hand encounter did in fact take place, when it didn't.

CONSEQUENTIAL AND DECISION VALUE AND THE IRRELEVANCE OF IRRELEVANT
ALTERNATIVES IN CHOICE PROBLEMS

Abstract: In this paper I propose a distinction between consequential and decision value in the rational assessment of choice. By making this distinction I provide an alternative taxonomy for dealing with rational choice problems. I introduce the notion of decision value, and show that it can explain a wider spectrum of choice problems than can be explained using consequential value alone. Perhaps more importantly, I show that decision value can make sense of choice problems that are considered irrational when assessed only by their consequential value. In doing so, I show that the consequentialist standard of rational choice, which I take to be the consensual means of addressing choice problems in economics and economically inspired philosophy, is too strong, as are the internal consistency standards of rationality that are closely related to it. The paper has two aims: to introduce the notion of decision value into the taxonomy by which choice problems are evaluated. And to demonstrate the implications that the notion of decision value has for the accepted normative standards of rational choice. In pursuing the latter aim, I will address the internal consistency conditions of rational choice as they are expressed in the principle of Independence of Irrelevant Alternatives (IIA). Using the notion of decision value, I propose a model of choice behavior which demonstrates that because the IIA only considers consequential (or "internal") value, and not decision value, it is too strong as a standard for rational choice.

1. INTRODUCTION

In this paper I propose a distinction between consequential and decision value in the rational assessment of choice. By making this distinction I provide an alternative taxonomy for dealing with rational choice problems. I introduce the notion of decision value, and show that it can explain a wider spectrum of choice problems than can be explained using consequential value alone. Perhaps more importantly, I show that decision value can make sense of choice problems that are considered irrational when assessed only by their consequential value. In doing so, I show that the consequentialist standard of rational choice, which I take to be the consensual means of addressing choice problems, is too strong, as are the internal consistency standards of rationality that are closely related to it.

I see this paper as having two aims. The first is to introduce the notion of decision value into the taxonomy by which choice problems are evaluated. The second is to demonstrate the implications that the notion of decision value has for the accepted normative standards of rational choice. In pursuing the latter aim, I will address the internal consistency conditions of rational choice as they are expressed in the principle of Independence of Irrelevant Alternatives (IIA). Using the notion of decision value, I propose a model of choice behavior which demonstrates that because the IIA only considers consequential (or “internal”) value, and not decision value, it is too strong as a standard for rational choice.

The paper will proceed as follows. I begin by introducing the notion of decision value. I then present the principle of Independence of Irrelevant Alternatives (IIA) and show how

alternatives that are independent and thus irrelevant in terms of consequential value are not necessarily irrelevant in terms of decision value. Thereafter I present a well-known counterexample to the IIA put forward by Amartya Sen (Sen 1993). I show that the force of Sen's counterexample can be undermined. All the same, I relate to the notion of epistemic value which Sen introduces, and argue that its scope is wider than Sen's example suggests. I then propose that decision value is closely related to Sen's notion of epistemic value but that it doesn't encounter the same difficulties. I show that decision value succeeds where Sen's epistemic value fails. Constructively, I propose that an adequate model for rational choice amounts to the conjunction of consequential and decision-sensitive considerations, and suggest that this amounts to what I see as falling under the notion of decision value. Decision value, as I envision it, is the weighted sum of the consequential and decision-sensitive utilities of an alternative in a choice set. As such I believe that it provides a more adequate and realistic standard for the evaluation of choice.

2. CONSEQUENTIAL AND DECISION VALUE

When it comes to questions of choice, consequentialist thinking considers the consequential value – typically tangible losses and gains - of alternatives in a choice set. On a consequentialist understanding, a rational choice – by which I understand a choice that is normatively commendable, is a choice that maximizes tangible value among alternatives. If for instance you make a rational choice between cars, the car you choose is purportedly the one with the greatest consequent value out of all the alternatives on a particular tangible dimension. The chosen car is in this sense the car that maximizes

value in relation to the available alternatives. Choice, in this sense, is governed by an assessment of the value to which each of the alternatives amounts to according to a particular decision rule. Therefore if a group of cars is assessed according to comfort, the most comfortable car will have the greatest consequent value, just as the most spacious car will have the greatest consequent value in terms of space. This position is, in brief and very generally, what I take to be the paradigmatic way of evaluating rational choice.

2.1 A CALL FOR A NEW PARADIGM

No doubt this staging of what I take to be the paradigm of rational choice is concise and simplified. Yet for my purpose at present the description is adequate. There does after all seem to me to be something rather simplistic in the approach underlying the paradigm I wish to discuss. And in addition to its simplistic appearance, there also seems to be something descriptively inadequate about it. As opposed to what the consequentialist standard of choice suggests, rationality appears to involve more than just choosing according to tangible gains and losses¹⁰⁰.

¹⁰⁰ Thinking about decision making in a non-consequential manner is not new in the social sciences. A field in which it is often discussed is the field of voting. See for example the recent paper by Harel and Shayo (Shayo and Harel 2012) on non-consequential voting. Others used the term "expressive decision making" or in the context of voting "expressive voting" (see (Carter and Guerette 1992)). I am grateful to Eyal Winter for drawing my attention to this matter. Additionally, there is a philosophical tradition that opposed consequentialism. See for instance works by: Nick Baigent, Wulf Gaertner, Prasanta Pattanaik, Yongsheng Xu and Kotaro Suzumura. Moreover, an important objection to consequentialism can be found in deontological theories of ethics that oppose the consequentialist assumption that choices – acts and/or intentions – are to be morally assessed solely by the states of affairs they bring about. Nonetheless, because

A rational choice is typically more than just a matter of maximizing consequential value; it often involves considering things from a long-term vantage point and not only according to immediate gain or causal utility considerations, such as which choice will lead to such and such an end. A rational choice is normally also a consistent choice that accords with what one would be likely to do in similar circumstances. It is also regarded as a coherent choice that relates to what one would be expected to do in relation to one's existing principles and beliefs. It will often also have symbolic and expressive value that amounts to more than its tangible or causal utility. But the consequentialist approach to rational choice doesn't appear to take such factors into consideration and this is partially why I consider it inadequate as a normative and descriptive standard for the evaluation of choice.

Decision value, as I shall refer to it here, augments the consideration of consequential value in choice problems. It affords what appears to be a more adequate evaluative standard for rational choice, while also providing a more realistic description of it. Broadly speaking, decision value describes the value that a choice has in and of itself, aside from the tangible value to which it leads. And as I shall demonstrate, this is an important factor to consider.

I first came across the notion of decision value in Robert Nozick's *The Nature of Rationality* (Nozick 1994) where it is presented as part of a broader project that is aimed at expanding traditional utility theory. The notion of decision value and the framework of rationality within which Nozick locates it seem to me to be enormously important for

this paper is focused on economic conceptions of choice, the philosophical literature will not be considered in the present context.

choice and decision theory, and I am quite surprised that its importance has not been duly recognized in the literature. Setting the details of Nozick's discussion aside, much of what I will propose with regard to decision value in what follows has its roots in Nozick's account, and more specifically in his discussions of principles (Nozick 1994, sec. I) and the implications of decision value for the prisoner's dilemma (Nozick 1994, 50–59). I will import several of the points which Nozick emphasizes into my own conception of decision value.

Nozick's discussion in *The Nature of Rationality* focuses on the inadequacies of decision theory as a description model of what actually occurs. Nozick's model offers a broader model of decision theory and a more overall view of rationality than suggested by alternative approaches. Nozick moves away from the simplistic view of rationality being exclusively concerned with consequentialist thinking to highlight the importance of the symbolic utility of actions that can stand for or symbolize different things relevant to the assessment of choice. Nozick suggests that we should talk about maximizing decision-value rather than merely maximizing utility. For Nozick, causally expected utility is that which relates means to ends. And he stresses that this is merely one component of rationality. My own conception of decision value builds on Nozick's account and expounds on it in suggesting that it can explain otherwise irrational choices as rational.

Here is my own definition of decision value:

- Decision value is the weight that is granted to the act of choosing a particular alternative, aside from the weight given to that alternative in lieu of any further outcome or consequence to which it leads; it is the value which choosing *that*

alternative will have in relation to one's long term view of things and in relation to one's principles, beliefs and commitments. It is what the choice of a particular alternative *means* – what it can come to symbolize; what it expresses, to oneself and to others.

In choice problems, considering the decision value of alternatives is an additional prism by which to evaluate choice. Choices often symbolize or stand for something in addition to the tangible value which they offer. Because of this, the utility of an alternative in a choice set can be imputed back not just along a causal, or consequential connection (e.g., eating meat to meet the goal of satisfying hunger) but also along a symbolic connection (e.g., what eating meat symbolizes for the person who chooses to eat it – think for instance of someone who was once a vegetarian).

My goal in introducing the notion of decision value is to distinguish two different kinds of value that I believe *ought* to be considered in the evaluation of rational choice. As noted, I think that the prevalent standard of rational choice only considers the consequential value of an alternative. In the present context I wish to augment the traditional evaluative toolkit by introducing an additional standard. As I shall demonstrate, this addition can explain choice-behavior that when only regarded through the prism of consequential value, can erroneously be considered irrational.

2.2 DECISION VALUE IN ACTION

Decision value should actually be very familiar. It is evident both in real-life situations of choice and in folk and commonly accepted conceptions of rationality. In real-life situations there are numerous considerations that lead one to make choices, not all of

which one is aware of. But within the scope of the considerations that people are aware of, some choices have a value in and of themselves – in the sense that they are significant beyond the value that they hold as alternatives. Promises are a good example. If I have made a promise to John and you offer me something of value in exchange for telling you his secret, in contemplating your offer I will typically consider not only what damage this would do to John and to my relationship with him (the causal or consequential value of perusing this alternative) but also what this would *mean* for my ability (and reputation) for keeping (and making) promises in the future (the decision value of my choice). Consider the following case as an illustration.

Ben is out for lunch with his friend Pat. The waiter is ready to take their order and, noticing Ben's Yamaka, announces that as of today the restaurant has a kosher kitchen. Ben keeps kosher and normally only eats meat at home, because he can never be sure how kosher other kitchens are. Ben comes to this restaurant quite frequently. But because of his religious dietary concerns he always orders a salad or a vegetable dish, and never a meat dish. But now that he has been told that the kitchen is kosher, he can, supposedly, choose anything that he fancies from the menu. That is, he can maximize tangible gain amongst the available alternatives, choosing the dish that he fancies instead of choosing according to religious dietary restrictions. In this sense because his religious restrictions are no longer of concern (and we assume that Ben has no reason not to think that the kitchen is not kosher, as the waiter says it is), he should choose the dish that he desires most.

Yet if Ben chooses *only* according to what he fancies - according to the tangible utility of the dishes on the menu, he will be deviating from a rule or perhaps a principle that he has

always held in reserve – an entrenched rule, perhaps one that his father conveyed to him when he was a boy – something along the lines of “Ben, make sure you only eat meat at home.” Because of this, eating meat outside of his own home is a symbolically significant choice for Ben, whatever the reasons that he happens to have in favor of doing so. If he has meat now he may begin eating out more often, and this may lead him to weaken his kosher standards altogether, something of which he is extremely fearful. This fear is part of what eating meat not-at-home symbolizes for Ben.

If this scenario does not seem farfetched, and the reader deems that it is plausible that issues such as these do concern Ben when he is considering the restaurant menu, then we can conclude that whatever Ben ends up choosing from the menu, his choice is not only a function of the consequential value of the dish that he chooses but also a function of its significance in relation to a broader framework of rules, principles, and concerns. His choice of dish from the menu will be based both on the consequential and on the symbolic dimensions that comprises that alternative’s decision value. Hence Ben’s abstaining from eating meat can be expressive and serves as an example of non-consequential reasoning.

3. CONSEQUENTIALISM AND CONSISTENCY

There is a widespread belief according to which rational choice is consistent choice, and that the nature of this consistency is *internal* consistency. Generally speaking, internal consistency is thought of as the type of consistency that can be established ‘within’ a choice set, among the alternatives that comprise it. On the internal consistency conception, whether a choice is consistent is a matter which depends on the alignment of

the alternatives chosen. Consistency, in this sense, depends on the alternatives chosen and not on the *reasons why* they were chosen. On this understanding, consistency does not invoke anything relating to mental considerations or decision rules by which choices are typically made (Hicks 1986; Sen 1993), nor does it appeal to any values that may broadly speaking reside inside a person's head (Hicks 1986). Just like the consequentialist approach to choice, the internal consistency conditions of rationality do not relate to what a choice *means* for the person choosing, but only to what it tangibly amounts to. It is because of this affinity between the consequentialist approach to choice and the internal consistency conditions of rationality that I have introduced the topic here. Furthermore, much of what I say with regard to the insufficiency of consequential value will apply to the internal consistency conditions of rationality.

3.1 A FOUNDATIONAL PROBLEM WITH INTERNAL CONSISTENCY

Aside from suffering from the same descriptive shortfall and as consequentialism, the internal consistency model of rationality is problematic in a foundational sense (Sen 1993). The foundational problem is that it isn't clear in what sense choices *can* be compared for consistency. Statements such as P and \neg P can be compared for consistency, but choices are not like statements in the sense that they do not have an inherent comparable logical structure. More specifically, the statements P and \neg P are *in themselves* inconsistent. Whereas the two choices – x and y, from {x,y} and {x,y,z}, do not, in themselves, appear to entail any such inconsistency. We might suppose that an inconsistency of choices may be a matter of different choices not serving what we assume to be their common goal. Thus if we had some information about what the person choosing x from {x,y} and y from {x,y,z} *was trying to do* – i.e., if he is a maximizer, *what*

he is trying to maximize, we might be able to infer statements adequately describing each of the two choices made, and we would thus be able to see whether these statements conflict (Sen 1993). Yet to gather what the person was trying to do would be to attain information beyond that which is available ‘internally,’ in the alternatives x and y themselves. And therefore at a minimum, there does not appear to be any immediately self-evident reason to think that inconsistency among choices is like the independent and self-standing inconsistency between statements. For one thing, it isn’t clear *how* or *which* choices negate each other. What’s more, it isn’t clear that it is possible to regard choices as conflicting without relating to the objectives that those choices serve.

Hence if the basis for an evaluation of rational choice is consistency, it isn’t clear how different choices can violate consistency. The internal consistency condition of rationality is problematic because it leaves out the role of the mental considerations – the purpose, the goal, the function - of the person making the choices. In so doing it is extremely similar to the approach that only considers consequential value without considering decision value.

4. THE IIA

Having identified the affinity between the consequentialist approach to choice and the internal consistency conditions of rationality, I will now demonstrate how decision value exposes both of these as insufficient for dealing with non-consequential decision-sensitive features in choice problems. I will do so by addressing the IIA, a paradigmatic model for rational choice.

There are various versions by which the principle of Independence of Irrelevant Alternatives (IIA) can and has been understood, but in the present context I will focus only on the following version, related to by Amartya Sen¹⁰¹. Other versions of the IIA are not important for my present purpose, which is limited to making a claim about the IIA construed as follows¹⁰²:

- If you choose a particular alternative from S (some choice set) and the alternative that you choose belongs to a subset T of S , you must choose that same alternative from T as well: $[[x \in C(S) \ \& \ x \in T \subseteq S] \Rightarrow x \in C(T)]$.

Choosing this way is choosing consistently. Moreover, a violation of the IIA thus construed – i.e., inconsistency between your choice in S and your choice in T - will look something like the conjunction of the following pair of choices:

- Choice #1: $\{y\} = C(\{x, y, z\})$; y is chosen out of the choice set $S = \{x, y, z\}$
- Choice #2: $\{x\} = C(\{x, y\})$; x is chosen out of the choice set $T = \{x, y\}$

In this case, the reason Choice #1 and Choice #2 are inconsistent is because a person who chooses y (and rejects x) in Choice #1 goes on to choose x (and reject y) in Choice #2. The internal consistency conditions of the IIA imply that if a person modifies (or reverses) his choice from Choice #1 to Choice #2, then:

¹⁰¹ Sen refers to the IIA in his discussion of ‘contraction consistency’. And while it seems that IIA belongs to contraction consistency, for our purposes we shall refer only to the IIA. The IIA appears to have been introduced by John Nash in (Nash 1950). A related idea appears in Samuelson’s (Samuelson 1938) paper on the Axiom of Revealed Preference.

¹⁰² For an overview, and clarification, of the various formulations of the IIA, see: (Ray 1973).

- a) *The choice* he has made is not rational.
- b) *He* is not choosing rationally.

The idea here is that if the alternatives in each of the choice sets are considered by their consequential value, then the presence of alternative z should not influence whether x or y is chosen out of a set which includes x and y and where, consequentially, the choice is seemingly only between these two alternatives. If z is relevant to the choice set, it is relevant to the consequential assessment of the choice set as a whole. But if the choice that is made in both Choice #1 and in Choice #2 is between x or y , as exhibited in the actual choices made in each case – y and then x accordingly, then z is consequentially irrelevant between x and y , and therefore a reversal of preference exhibited by the choice of y in S and x in T, is irrational.

Because of how these choices are evaluated according to the standards of the IIA, I would like to suggest that even though z might be considered irrelevant when Choice #1 is evaluated by consequential value [CV] alone, it needn't necessarily be irrelevant when Choice #1 is evaluated by the consequential value of z *in conjunction* with its other decision-sensitive attributes, which together amount to its decision value [DV]. Hence I would like to suggest that the decision value of z may provide an alternative evaluative framework for the difference in choice (or reversal of preference) between S and T. On this suggestion, to be explored in the following sections, z may be evaluated in a manner that modifies the way in which x and y are assessed in relation to S and T.

5. AMARTYA SEN AND THE EPISTEMIC VALUE OF IRRELEVANT ALTERNATIVES

In the context of his discussion of the internal consistency conditions of rationality, Amartya Sen argues that in addition to their ordinal or consequential value¹⁰³, alternatives in a choice set can have *epistemic value* (Sen 1993). On this suggestion, the addition of an otherwise independent alternative to a choice set can modify the epistemic value of other alternatives in that set, and can lead to what may ostensibly seem like an exhibited preference reversal or inconsistency between choices. Sen argues that once epistemic value is recognized as a critical component of rational choice, the internal consistency conditions of rationality, which only consider the consequential value of alternatives in choice sets, is shown to be inadequate and deficient.

Here is how Sen illuminates the role of epistemic value in the normative evaluation of choice:

¹⁰³ Sen does not refer to consequential value but I interpret his reference to ordinal value as being at root the same as consequential value. Sen's discussion of the IIA is part of his broader analysis of the internal consistency of choice (Sen 1993) examined under various consistency conditions, one of which is 'Contraction Consistency' – what he also refers to as "property α (alpha)," or the IIA. Sen's discussion of the internal consistency model for rational choice is motivated by two concerns which he expresses at the outset of his discussion (Sen 1993). The first of these questions whether the consistency condition of rationality must take the form of *internal* consistency of choice. The second questions whether it is possible to establish choice consistency in a *context-independent* way. Sen emphasizes at the outset of his discussion that internal consistency is unproblematic insofar as internal inconsistency is *implied* by a given utility function. His critique is instead aimed at the *a priori imposition* of such internal consistency conditions (Sen 1993, 496) as conditions for rational choice.

Given a choice between having tea at a distant acquaintance's home (x), and not going there (y), a person who chooses to have tea (x), may nevertheless choose to go away (y), if offered – by that acquaintance – a choice of having tea (x), going away (y), and having some cocaine (z) (Sen 1993, 502).

What Sen seems to have in mind in this context is that the epistemic value of an alternative in a choice set can provide relevant decision-sensitive information. In this case alternative z provides information that is relevant for the evaluation of x and y . In Sen's example, option z – 'having some cocaine,' conceivably has an epistemic value in addition to its consequential value in the choice set; z , the alternative of 'having some cocaine,' almost certainly has a meaning beyond its tangible meaning. It seems to tell us something about our 'distant acquaintance,' which may consequently affect our choice of what to do with regard to x or y – both of which are alternatives that in themselves are independent of z but which are nonetheless influenced by what we learn from z . Choice z can, in this sense, provide us with information about the person offering us options x and y and this information may "quite reasonably" (Sen 1993) affect the preferential ranking of choices x and y .¹⁰⁴ And this demonstrates that it is quite plausible for consequentially irrelevant alternatives to have epistemic value which makes them epistemically relevant in situations of choice.

¹⁰⁴ In fact, the information provided to us by z – 'having some cocaine' – may change the choice set altogether. Having been offered option z by our distant acquaintance, options x and y are perhaps no longer *the same* options that we had in the previous choice set. They are, conceivably, *new* options given to us by someone who also offered us cocaine. I will discuss this possibility in the next section.

According to Sen, because alternatives in situations of choice can have an epistemic value in addition to their consequential value, acts of choice which violate the IIA on the consequentialist standard and are thus judged inconsistent and hence also irrational, may no longer be appraised in this way when they are considered in the appropriate deliberative setting. Once the epistemic value of alternatives is taken into account, otherwise inconsistent behavior can be made intelligible. It can be made intelligible because incremental information is provided as a means of explaining an otherwise irrational choice. On this understanding it seems that the internal consistency conditions of the IIA are too strong. They are too strong because they can be violated by choices that, once explained in reference to epistemic value, may seem entirely rational.

6. A DIFFICULTY WITH EPISTEMIC VALUE CONSIDERATIONS

While it is intuitively appealing, Sen's notion of epistemic value is not immune to problems. And neither is his critique of the internalist standard of rational choice. If we accept Sen's argumentation, the epistemic value of an alternative in a choice set can transform the significance of other alternatives in the choice set, as the offer of cocaine z can transform the significance of alternatives x and y in the example related to above.

The main problem with Sen's proposal is that if different choice sets are made up of *different* alternatives, there doesn't appear to be a problem of internal inconsistency, as the IIA advocates. If an alternative z has an epistemic value that *really* transforms the nature of the other alternatives – x and y - in the choice set, then these alternatives are in fact no longer *the same* alternatives. And neither are they factors with regards to which consistency can be assessed. Given the incremental decision-sensitive information

afforded by the epistemic value of z , x and y become \hat{x} and \hat{y} - which symbolize x and y albeit with an incremental epistemic infusion of decision-relevant information. And there is hence no longer a problem of inconsistency between Choice #1 and Choice #2.

To make this clear let me illustrate this finding in relation to Sen's own counterexample. When your distant acquaintance offers you cocaine, the offer does not so much impact the way in which you *perceive* the alternatives on offer – e.g., 'having tea at your acquaintance's home' (x) or 'not going there' (y). Instead, the proffering of illegal indulgences actually *changes* the available alternatives set before you.

On this reading, the two seemingly common options in the two choice sets – Choice #1: {*tea at your acquaintance's house* (x), *not going there* (y)} and Choice #2: {*tea at your acquaintance's house* (x), *not going there* (y), *cocaine* (z)} are not *actually* the same options. Having been offered cocaine by your distant acquaintance, the *tea at your acquaintance's house* and *staying at home* in Choice #2, turn into something along the lines of *tea-offered-by-someone-who-offered-you-cocaine* (\hat{x}) and *staying-at-home-instead-of-going-to-have-tea-with-someone-who-offered-you-cocaine* (\hat{y}). And because these alternatives are in fact different from those offered in Choice #1, there is no inconsistency in the exhibited preference reversal from Choice #1 to Choice #2. What we are seeing are simply two choices from two *different* sets of alternatives.

In a recent paper, Michael Neumann (Neumann 2007) suggests that in Sen's example, the *identity* of the alternatives in these choice sets *changes* once epistemically relevant information is provided by the fact that the 'distant acquaintance' has offered cocaine (z). On Neumann's interpretation, each alternative in a choice set has an *identity* that includes

all its decision-sensitive features. And once these features include information such as that the person who is offering tea is also a person who offers cocaine, the identity of these alternatives is no longer the same.

6.1 WHERE DOES THIS LEAVE US?

If we accept this criticism of Sen's notion of epistemic value we may have to give up the idea that the epistemic value of alternatives in choice sets is in fact a significant factor for the normative evaluation of choice. In this case we would seemingly have to recede back to the internal consistency conditions of rationality and its consequentialist roots of choice evaluation. Nonetheless, if we can show that epistemic value extends further than Sen's example suggests, then the problems incurred by that example may not apply to epistemic value in general, but only to the type of epistemic value exhibited in Sen's counterexample, that which pertains to the provision of incremental decision-sensitive information.

There are, I contend, other types of epistemic value which do not face the same problems faced by Sen's example. Speaking generally, I think that epistemic value is broader in scope than what is involved in the provision of incremental information. And I think that the notion is under-described in Sen's account. I contend that once a more directed criticism of the consequentialist standard of rational choice is set against the internal consistency conditions of the IIA, a more convincing model of choice can be provided. This will be my aim in the subsequent section of this paper.

6.2 EPISTEMIC VALUE AND DECISION VALUE

Before I proceed to show how the notion of decision value can be useful, I will propose my own working definition of what I mean by epistemic value, to which decision value is closely related. I believe the differences between decision value and epistemic value are subtle ones, but regrettably I cannot at present provide a satisfactory demarcation of the differences between them.

- In choice sets, alternatives can have a broader significance than how they measure on a single dimension, according to their consequential value. This broader significance includes what the alternatives symbolize and how they integrate with the other beliefs, intentions, objectives, and principles that a person has. It is the overall significance of an alternative that comprises its epistemic value, to which decision value is closely related.

As we saw, the epistemic value of alternatives in choice sets can include the information that the alternatives provide the person choosing in relation to the available alternatives, in the way z provides decision-relevant information that pertains to x and y . But the epistemic value of alternatives may also include how *familiar* the alternatives are, what they *represent*, and what choosing them *means*. The latter comprise what I have referred to as the decision value of alternatives. Decision value comprises decision-sensitive features that may have a bearing on the way in which alternatives are assessed in choice conditions.

My aim in the final section of this paper is to demonstrate how decision value can explain choices that are consequentially inconsistent. I propose an evaluative model that

incorporates decision value in the assessment of choice. On the basis of this model I show that the IIA and the consequentialist standard of rationality are too strong. I then provide an example that fits the evaluative model that I propose. This example would purportedly be considered irrational on the consequential and internal consistency model of rational choice but seems perfectly plausible on the decision value model of rational choice.

7. THE EPISTEMIC RELEVANCE OF CONSEQUENTIALLY IRRELEVANT ALTERNATIVES

The following is the model of decision value that I propose. It provides an augmented standard of evaluation of alternatives in choice sets:

- $\{x, S\} > \{y, S\}$; when x is considered in the context of S ($\{x, y, z\}$) it has a greater value than y has when it is considered in the same context.
- $\{y, T\} > \{x, T\}$; when y is considered in the context of T ($\{x, y\}$) it has a greater value than x has when it is considered in the same context.

As outlined above, S and T are two different choice sets. T is a subset of S , but it is nonetheless a different set. What I am proposing in this model is that there are cases in which x and y can have different values depending on the set to which they belong. Ultimately, the set to which x or y belong determines the value of each alternative in accordance with the conjunction of the consequential *and* decision-sensitive features that belong to it. Here is an illustration of how this can happen.

7.1 A CASE OF DECISION VALUE

You are offered an apple $\{x\}$ and an ice cream $\{y\}$. You choose the apple $\{x\} = T = C(\{x, y\})$, perhaps because it is healthier. You are then also offered a cigarette (z). In this case you choose the ice cream $\{y\} = S = C(\{x, y, z\})$, perhaps because you feel that since you refused the cigarette you deserve an ice cream.

Note that this narrative adheres to my decision value model - $\{x, T\} > \{y, T\}$ but $\{y, S\} > \{x, S\}$; x has a higher value than y in T , and Y has a higher value than x in S . In this example, the *value* of each alternative (denoted by the ' $>$ ' symbol) is a function of decision value. It is the conjunction of consequential and decision-sensitive features of *that* alternative. Here is a narrative that explains why this is so.

When an {apple (x) and an ice cream (y)} are on offer, the apple has a higher consequential value than it does when an {apple (x), an ice cream (y), and a cigarette (z)} are on offer. The fact that the cigarette is on offer may introduce a new preferential dimension on the basis of which the three alternatives x , y , z , are assessed. Here is a possible explanation for such a change in preferential dimension. Because a cigarette (z) is the type of thing the choice of which (that is, smoking) can be significant in and of itself – such as its swift induction into habitual smoker-behavior, its representation of a yield to temptation - it can elicit non non-consequential features of the alternatives to which it relates. Because of this, the apple (x), the ice cream (y), and the cigarette (z) are assessed by their overall decision value. And since we are told that it was chosen, we can assume that ice cream (y) has the greatest decision value in relation to x and y . Why does ice cream (y) have the greatest decision value? Perhaps because it is a conciliatory

alternative that results from the tension between the temptation of a cigarette (z) and the symbolic cost of yielding to it, together with the level of desirability of the apple (x). But this is only a speculative guess. Ultimately, the reason why ice cream was chosen is because it is the preferred choice in terms of the consequential and decision-sensitive features of the remaining alternatives; otherwise put: it is chosen because it is the choice with the greatest overall decision value.

8. IS DECISION VALUE NOT SIMPLY A CONCLUSIVE FORM OF CONSEQUENTIAL VALUE?

There is a temptation, often put forward by economists, to suggest that what I have here referred to as decision value can actually be understood as being part of consequential value, something that accordingly makes the notion of decision value superfluous¹⁰⁵. The idea is that any and all decision-sensitive considerations that relate to an alternative in a choice set ought to be included *as part* of that alternative's consequential value. The idea is that if breaking a promise has value with regards to a person's ability to keep promises in the future, this feature is an inherent part of the consequential evaluation of breaking a

¹⁰⁵ This goes back to an earlier debate about the extensional equivalence of act and rule utilitarianism (Neumann 2007, 81). Setting the details of the debate aside, the standard that emerged from this debate was that the right way to describe an alternative in a choice set was to include in its description all its causally consequential properties (for relevant reading on this debate see: (Ezorsky 1968; Lyons and Lyons 1965; Nozick 1994; Rescher 1969)). The goal was to incorporate all the ancillary values that an alternative is expected to have *if* chosen. The idea is that such properties “must be included in the description because [they] effect the act's utility and therefore – in a utilitarian context – the value of the act” (Neumann 2007, 81).

promise. On this understanding, the overall significance of an alternative, in addition to the tangible value of choosing it, is part of that alternative's consequential utility.

Here is my brief response to this suggestion. The dismissal of decision value in view of the claim that it is part of consequential utility gives rise to an over-simplified picture of choice that is distortive both in the way it describes how choices are made and in the normative framework it proposes for understanding choices. The notion of decision value that I have introduced here enables a more realistic and familiar conception of rationality for which consequential utility is insufficient.

9. CONCLUSION

I proposed an augmented taxonomy for the evaluation of rational choice based on the introduction of the notion of decision value. Decision value highlights the value that a choice of an alternative can have in and of itself. I argued that alternatives in choice sets ought to be assessed not only by their consequential value but also by their decision value, which includes consequential value as well as the non-consequential significance that an alternative may have. I proposed that the decision value of an alternative in a choice set is the weighted sum of its consequential value and non-consequential decision-sensitive features. I suggested that the decision value model of choice provides a more realistic description and an improved normative standard for the evaluation of choice.

BIBLIOGRAPHY

- Ajzen, Icek. 1977. "Intuitive Theories of Events and the Effects of Base-rate Information on Prediction." *Journal of Personality and Social Psychology* 35 (5): 303–314.
- Anderson, John R., and Robert Milson. 1989. "Human Memory: An Adaptive Perspective." *Psychological Review* 96 (4): 703–719.
- Armstrong, D. M. 1973. *Belief, Truth and Knowledge*. 1st ed. Cambridge University Press.
- Aumann, Robert J. 1976. "Agreeing to Disagree." *The Annals of Statistics* 4 (6) (November 1): 1236–1239.
- . 2008. *Rule-Rationality Versus Act-Rationality*. Discussion Paper Series. Center for Rationality and Interactive Decision Theory, Hebrew University, Jerusalem.
<http://ideas.repec.org/p/huj/disap/dp497.html>.
- Barbey, Aron K., and Steven A Sloman. 2007. "Base-rate Respect: From Ecological Rationality to Dual Processes." *The Behavioral and Brain Sciences* 30 (3) (June): 241–254; discussion 255–297.
- Bar-Hillel, Maya. 1980. "The Base-rate Fallacy in Probability Judgments." *Acta Psychologica* 44 (3) (May): 211–233.
- . 1982. "Ideal Evidence, Relevance and Second-order Probabilities." *Erkenntnis* 17 (3): 273–290.
- Bar-Hillel, Maya, and Baruch Fischhoff. 1981. *When Do Base Rates Affect Predictions?* *Journal of Personality and Social Psychology*, Vol 41(4), Oct 1981, 671-680.
- Bar-Hillel, Maya, and Efrat Neter. 1993. "How Alike Is It Versus How Likely Is It: A Disjunction Fallacy in Probability Judgments." *Journal of Personality and Social Psychology* 65 (6): 1119–1131.
- Bermúdez, José Luis, and Sebastian Gardner. 2003. *Art and Morality*. Psychology Press.
- Bickel, J. Eric. 2007. "Some Comparisons Among Quadratic, Spherical, and Logarithmic Scoring Rules." *Decision Analysis* 4 (2) (June 1): 49–65.
- Binkley, Timothy. 1977. "Piece: Contra Aesthetics." *The Journal of Aesthetics and Art Criticism* 35 (3) (April 1): 265–277.

- Brier, Glenn W. 1950. "Verification of forecasts expressed in terms of probability." *Monthly Weather Review* 78 (1) (January): 1–3.
- Budd, Malcolm. 2003. "The Acquaintance Principle." *The British Journal of Aesthetics* 43 (4) (October 1): 386–392.
- Carroll, Noël. 1999. *Philosophy of Art: A Contemporary Introduction*. annotated ed. Routledge.
- . 2004. "Non-perceptual Aesthetic Properties: Comments for James Shelley." *British Journal of Aesthetics* 44 (4): 413–423.
- Carter, John R., and Stephen D. Guerette. 1992. "An Experimental Study of Expressive Voting." *Public Choice* 73 (3) (April): 251–260.
- Christensen, David. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 116 (2) (April): 187–217.
- . 2004. *Putting Logic in Its Place: Formal Constraints on Rational Belief*. Clarendon Press.
- . 2010. "Higher-order Evidence," *Philosophy and Phenomenological Research*, 81: 185–215.
- . 2011. "Disagreement, Question-Begging, and Epistemic Self-Criticism." *Philosophers' Imprint* 11 (6): unknown.
- Cohen, Stewart. 2002. "Basic Knowledge and the Problem of Easy Knowledge," *Philosophy and Phenomenological Research*, 65, (2,): 309–329.
- Cowen, Tyler, and Robin Hanson. 2004. "Are Disagreements Honest?" Department of Economics George Mason University - Working Paper.
- Danto, Arthur. 1964. "The Artworld." *The Journal of Philosophy* 61 (19) (October 15): 571–584.
- David, Marian. 2001. "Truth as the Epistemic Goal." In *Knowledge, Truth, and Duty*, ed. M. Steup, 151–169. New York: Oxford University Press.
- Dégremont, Cédric, and Olivier Roy. 2009. "Agreement Theorems in Dynamic-Epistemic Logic." In *Logic, Rationality, and Interaction*, ed. Xiangdong He, John Horty, and Eric Pacuit, 5834:105–118. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dickie, George. 1974. *Art and the Aesthetic: An Institutional Analysis*. Cornell University Press.

- Diffey, T. J. 1982. "Aesthetic Instrumentalism." *The British Journal of Aesthetics* 22 (4): 337–349.
- Driver, Julia. 2006. "Autonomy and the Asymmetry Problem for Moral Expertise." *Philosophical Studies* 128 (3) (April): 619–644.
- Ehm, Werner. 2012. "Local Proper Scoring Rules of Order Two." *The Annals of Statistics* 40 (1) (February): 609–637.
- Elga, Adam. 2007a. "Reflection and Disagreement." *Noûs* 41 (3): 478–502.
 ———. 2007b. "How to Disagree About How to Disagree". In *Disagreement*, ed. Ted Warfield and Richard Feldman. Oxford University Press.
- Enoch, David. 2011. "Not Just a Truthometer: Taking Oneself Seriously (but Not Too Seriously) in Cases of Peer Disagreement." *Mind* 119 (476):953-997.
 ———. 2009. "How Is Moral Disagreement a Problem for Realism? ," *Journal of Ethics*, 13, (1).
- Ezorsky, Gertrude. 1968. "A Defense of Rule Utilitarianism Against David Lyons Who Insists on Tying It to Act Utilitarianism, Plus a Brand New Way of Checking Out General Utilitarian Properties." *The Journal of Philosophy* 65 (18): 533–544.
- Feldman, Richard. 2000. "The Ethics of Belief." *Philosophy and Phenomenological Research* 60 (3): 667–695.
 ———. 2005. "Respecting the Evidence." *Philosophical Perspectives* 19 (1) (December): 95–119.
 ———. 2006. "Epistemological Puzzles About Disagreement ," In *Epistemology Futures* ,, ed. Stephen Hetherington, 216–326 ,. Oxford University Press.
 ———. 2007. "Reasonable Religious Disagreements ," In *Philosophers Without Gods: Meditations on Atheism and the Secular*, ed. Louise Antony, 194–214. Oxford University Press.
- Feldman, Richard, and Ted A. Warfield. 2010. *Disagreement*. Oxford University Press.
- Foley, Richard. 1987. *The Theory of Epistemic Rationality*. First Pr. Thus. Harvard University Press.
- Foley, Richard. 1987. *The Theory of Epistemic Rationality*. First Pr. Thus. Harvard University Press.

- . 1993. *Working Without a Net: a Study of Egocentric Epistemology*. Oxford University Press US.
- Fricker, Elizabeth, and David E Cooper. 1987. "The Epistemology of Testimony." *Aristotelian Society Supplementary Volume* 61: 57 – 106.
- Gelfert, Axel. 2011. "Who Is an Epistemic Peer?" *Logos and Episteme* 2 (4): 507–514.
- Gendler, Tamar Szabo, and John Hawthorne. 2008. *Oxford Studies in Epistemology: Volume 2*. illustrated ed. Oxford University Press, USA.
- Gendler, Tamar Szabó, and John Hawthorne. 2008. *Oxford Studies in Epistemology*. Oxford University Press.
- Giacomo Bonanno, and Klaus Nehring. *Agreeing To Disagree: A Survey*. California Davis - Department of Economics. <http://ideas.repec.org/p/fth/caldec/97-18.html>.
- Gibbard, Allan. 2007. "Rational Credence and The Value of Truth." In: Gendler and Hawthorne (2008).
- Gigerenzer, Gerd, Wolfgang Hell, and Hartmut Blank. 1988. "Presentation and Content: The Use of Base Rates as a Continuous Variable." *Journal of Experimental Psychology: Human Perception and Performance* 14 (3) (August): 513–525.
- Gilovich, Thomas. 1993. *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. Free Press.
- Gneiting, Tilmann, and Adrian E Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–378.
- Goldberg, Sanford. 2009. "Experts, Semantic and Epistemic." *Noûs* 43 (4): 581–598.
- Hanfling, Oswald. 1992. *Philosophical Aesthetics: An Introduction*. Open University.
- Hansen, Robin, and Tyler Cowen. "Are Disagreements Honest?" *Journal of Economic Methodology* to appear. <http://hanson.gmu.edu/deceive.pdf>.
- Hanson, Robin. 2003. "For Bayesian Wannabes, Are Disagreements Not About Information?" *Theory and Decision* 54 (2): 105–123.

- Harman, Gilbert. 1973. *Thought*. Princeton University Press.
- . 1995. “Rationality.” In *Invitation to Cognitive Science*, ed. E. E. Smith and D. N. Osherson. MIT Press.
- . 1997. “Pragmatism and Reasons for Belief.” In *Realism/Antirealism and Epistemology*, ed. C. B. Kulp. Rowman and Littlefield.
- Hicks, J. R. 1986. *A Revision of Demand Theory*. Clarendon Press.
- Hopkins, Robert. 2000. “Beauty and Testimony.”
- . 2007. “Speaking Through Silence : Conceptual Art and Conversational Implicature.” In *Philosophy and Conceptual Art*, ed. Peter Goldie and Elisabeth Schellekens. Oxford University Press.
- . 2011. “How to Be a Pessimist About Aesthetic Testimony.” *Journal of Philosophy* 108 (3): 138–157.
- Horwitz, Paul. 2007. “Three Faces of Deference.” *Notre Dame Law Review*, Vol. 83, 2008.
- van Inwagen, Peter. 1996. “Is It Wrong Everywherhe, Always, and for Anyone to Believe Anything on Insufficient Evidence?” In *Faith, Freedom and Rationality*, ed. Jeff Jordan and Daniel Howard-Snyder, 137–154. Savage, Maryland: Rowman and Littlefield.
- John Geanakoplos, and Heracles M. Polemarchakis. 1982. *We Can't Disagree Forever*. Cowles Foundation for Research in Economics, Yale University.
- Jonathan A.K., Cave. 1983. “Learning to Agree.” *Economics Letters* 12 (2): 147–152.
- Jose, Victor. 2009. “A Characterization for the Spherical Scoring Rule.” *Theory and Decision* 66 (3): 263–281.
- Joseph Y. Halpern. 2000. *Characterizing the Common Prior Assumption*. EconWPA.
- Joyce, James. 2009. “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief.” In *Degrees of Belief*, ed. Franz Huber and Christoph Schmidt-Petri, 342:263–297.
- . 2007. “Epistemic Deference: The Case of Chance.” *Proceedings of the Aristotelian Society* 107 (2): 187–206.
- . 1998. “A Nonpragmatic Vindication of Probabilism.” *Philosophy of Science* 65 (4): 575–603.

- Kahneman, Daniel, and Amos Tversky. 1973. "On the Psychology of Prediction." *Psychological Review* 80 (4): 237–251.
- Kant, Immanuel. *The Critique of Judgement*. Forgotten Books.
- Kelly, Thomas. 2002. "The Rationality of Belief and Some Other Propositional Attitudes." *Philosophical Studies* 110 (2): 163–196.
- . 2003. "Epistemic Rationality as Instrumental Rationality: A Critique." *Philosophy and Phenomenological Research* 66 (3) (May): 612–640.
- . 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology, Volume 1*, ed. John Hawthorne and Tamar Gendler Szabo, 167–196. Oxford University Press.
- . 2008. "Disagreement, Dogmatism, and Belief Polarization," *The Journal of Philosophy*, 105, (10,): 611–633.
- . 2009. "Peer Disagreement and Higher Order Evidence," In *Disagreement*, ed. Ted Warfield and Richard Feldman. Oxford University Press.
- Keren, Arnon. 2007. "Epistemic Authority, Testimony and the Transmission of Knowledge†." *Episteme* 4 (3): 368–381.
- Kivy, Peter. 1991. "Science and Aesthetic Appreciation." *Midwest Studies In Philosophy* 16 (1) (September 1): 180–195.
- Konigsberg, Amir. *forthcoming*. "Epistemic Value and Epistemic Compromise, a reply to Moss." *Episteme*.
- . *forthcoming*. "The Problem with Ununiform Solutions to Peer Disagreement." *Theoria*.
- . 2012. "The Acquaintance Principle, Aesthetic Autonomy, and Aesthetic Appreciation." *British Journal of Aesthetics* 52 (2): 153-168.
- Kornblith, Hilary. 1993. "Epistemic Normativity." *Synthese* 94 (3): 357 – 376.
- Kripke, Saul A. 2011. "Two Paradoxes of Knowledge." In *Philosophical Troubles. Collected Papers Vol I*, ed. Saul A. Kripke. Oxford University Press.
- Kunda, Z. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3) (November): 480–498.
- Laetz, Brian. 2008. "A Modest Defense of Aesthetic Testimony." *Journal of Aesthetics and Art Criticism* 66 (4): 355–363.

- Lamarque, Peter, and Stein Haugom Olsen. 2004. *Aesthetics and the Philosophy of Art: The Analytic Tradition : an Anthology*. Wiley-Blackwell.
- List, Christian. working paper. “When to Defer to Supermajority Testimony — and When Not”. See: <http://eprints.lse.ac.uk/20074/>
- Livingston, Paisley. 2003. “On an Apparent Truism in Aesthetics.” *The British Journal of Aesthetics* 43 (3) (July 1): 260–278.
- Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence.” *Journal of Personality and Social Psychology* 37 (11) (November): 2098–2109.
- Lyons, Louis, and David Lyons. 1965. *Forms and Limits of Utilitarianism*. First ed. Oxford University Press.
- Margolis, Joseph. 1960. “Aesthetic Perception.” *The Journal of Aesthetics and Art Criticism* 19 (2) (December 1): 209–213.
- Matheson, Jonathan. 2009. “Conciliatory Views of Disagreement and Higher-Order Evidence.” *Episteme* 6 (October): 269–279.
- Meskin, Aaron. 2007. “Solving the Puzzle of Aesthetic Testimony.” In *Knowing Art*, ed. Matthew Kieran and Dominic McIver Lopes, 109–124. Dordrecht: Springer Netherlands.
- Milgrom, Paul, and Nancy Stokey. 1982. “Information, Trade and Common Knowledge.” *Journal of Economic Theory* 26 (1) (February): 17–27.
- Moses, Yoram, and Gal Nachum. 1990. “Agreeing to Disagree After All.” In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning About Knowledge*, 151–168. TARK '90. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Moss, Sarah. 2012. “Scoring Rules and Epistemic Compromise.” *Mind*. 120 (480):1053-1069.

- Nash, John. 1950. "The Bargaining Problem." *Econometrica* 18 (2): 155–162.
- Neumann, Michael. 2007. "Choosing and Describing: Sen and the Irrelevance of Independence Alternatives." *Theory and Decision* 63 (1) (April 17): 79–94.
- Nozick, Robert. 1994. *The Nature of Rationality*. New ed. Princeton University Press.
- Oddie, Graham. 1997. "Conditionalization, Cogency, and Cognitive Value." *British Journal for the Philosophy of Science* 48 (4): 533–541.
- Owens, David. 2000. *Reason Without Freedom: The Problem of Epistemic Normativity*. Routledge.
- Payne, John W., and Carnegie-Mellon University. 1976. *Cognition and Social Behavior*. Routledge.
- Percival, Philip. 2002. "Epistemic Consequentialism: Philip Percival." *Aristotelian Society Supplementary Volume* 76 (1): 121–151.
- Pillar, Christian [1]. 2001. "Normative Practical Reasoning: Christian Pillar." *Supplement to the Proceedings of The Aristotelian Society* 75 (July): 195–216.
- Ray, Paramesh. 1973. "Independence of Irrelevant Alternatives." *Econometrica* 41 (5): 987–991.
- Rescher, Nicholas. 1969. *Introduction to Value Theory*. University Press of America.
- Robin Hanson. 2003. "For Bayesian Wannabes, Are Disagreements Not About Information?" *Theory and Decision* 54 (2). *Theory and Decision*: 105–123.
- Robson, Jon. 2012. "Aesthetic Testimony." *Philosophy Compass* 7 (1) (January 1): 1–10.
- Rosen, Gideon. 2001. "Nominalism, Naturalism, Epistemic Relativism." *Nous* 35 (s15) (October): 69–91.

- Rubinstein, Ariel, and Asher Wolinsky. 1990. "On the Logic of 'Agreeing to Disagree' Type Results." *Journal of Economic Theory* 51 (1). *Journal of Economic Theory*: 184–193.
- Russell, Bertrand. 1996. *Unpopular Essays*. Psychology Press.
- Samuelson, P. A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica* 5 (17). New Series (February 1): 61–71.
- Schaper, Eva. 1987. *Pleasure, Preference and Value: Studies in Philosophical Aesthetics*. CUP Archive.
- Sen, Amartya. 1993. "Internal Consistency of Choice." *Econometrica* 61 (3) (May 1): 495–521.
- Shayo, Moses, and Alon Harel. 2012. "Non-consequentialist Voting." *Journal of Economic Behavior & Organization* 81 (1) (January): 299–313.
- Shelley, James. 2003. "The Problem of Non-Perceptual Art." *The British Journal of Aesthetics* 43 (4) (October 1): 363–378.
- Sibley, F. N. 1974. "The Inaugural Address: Particularity, Art and Evaluation." *Proceedings of the Aristotelian Society, Supplementary Volumes* 48 (January 1): 1–21.
- Sibley, Frank. 1959. "Aesthetic Concepts." *The Philosophical Review* 68 (4) (October 1): 421–450.
- . 1965. "Aesthetic and Nonaesthetic." *The Philosophical Review* 74 (2) (April 1): 135–159.
- Slovic, Paul. 2010. *The Feeling of Risk: New Perspectives on Risk Perception*. 1st ed. Routledge.
- Slovic, Paul, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor. 2007. "The Affect Heuristic." *European Journal of Operational Research* 177 (3) (March 16): 1333–1352.
- Slovic, Paul, Ellen Peters, Melissa L. Finucane, and Donald G. Macgregor. 2005. "Affect, Risk, and Decision Making." *Health Psychology: Official Journal of the Division*

- of Health Psychology, American Psychological Association* 24 (4 Suppl) (July): S35–40.
- Stecker, Robert. 1984. “Aesthetic Instrumentalism and Aesthetic Autonomy.” *The British Journal of Aesthetics* 24 (2) (March 20): 160–165.
- Steup, Matthias. 2001. *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press, USA.
- Thaler, Richard H., and Cass R. Sunstein. 2008a. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.
- Tormey, Alan. 1973. “Critical Judgments.” *Theoria* 39 (1-3): 35–49.
- Velleman, David. 2000. “On the Aim of Belief,” In *The Possibility of Practical Reason* , ed. David Velleman, 244–81 ., Oxford University Press.
- Webb, Mark Owen. 1993. “Why I Know About as Much as You: A Reply to Hardwig.” *The Journal of Philosophy* 90 (5) (May 1): 260–270.
- Wedgwood, Ralph. 2002. “The Aim of Belief.” *Philosophical Perspectives* 16 (s16): 267–97.
- Weitz, Morris. 1956. “The Role of Theory in Aesthetics.” *The Journal of Aesthetics and Art Criticism* 15 (1): 27–35.
- Welsh, Matthew B., and Daniel J. Navarro. “Seeing Is Believing: Priors, Trust, and Base Rate Neglect.” *Organizational Behavior and Human Decision Processes* (0). <http://www.sciencedirect.com/science/article/pii/S0749597812000489>.
- Werner Ehm, and Tilmann Gneiting. 2010. *Local Proper Scoring Rules*. Addendum to Technical Report no. 551.
- Whitcomb, Dennis. 2011. *Social Epistemology: Essential Readings*. Oxford University Press.

White, Roger. 2005. "Epistemic Permissiveness." *Philosophical Perspectives* 19 (1): 445–459.

Wollheim, Richard. 1980. *Art and Its Objects*. 2nd ed. Cambridge University Press.

Zemach, Eddy M. 1997. *Real Beauty*. Pennsylvania State University Press.

תקציר בעברית

חיבור זה עוסק במספר סוגיות פילוסופיות שקורמות עור וגידים מתוך השיקולים הנורמטיביים הכרוכים בבחינה ביקורתית של האמנות. החיבור מתמקד בקבלת החלטות בתנאים של אי ודאות, בהינתן סוגי מידע שונים – מידע שבמקרים רבים הוא מוגבל, מעורפל, ודו משמעי. החיבור מורכב מארבעה שערים. בשער הראשון שני חלקים העוסקים במשמעותה האפיסטמית של אי הסכמה וביחסה לפשרה אפיסטמית. שני חלקיו של השער מהווים תרומה חדשה לדיון הפילוסופי העכשווי העוסק בסוגיות אלה. השער השני של החיבור עוסק בתפקיד שממלא מידע הסתברותי מקדים במסגרת שיפטים אפיסטמיים אודות מידע חדש. הדיון מציע ניתוח חדשני לתפקידם של שיפטים אודות רלבנטיות – 'שיפוטי רלוונטיות' – של מידע בסוגיות של בחינת מידע וקבלת החלטות. בשער השלישי שני חלקים העוסקים בתמות מתחומי האפיסטמולוגיה והאסתטיקה. כל חלק בוחן שאלות נורמטיביות שונות על אודות שיפטים אסתטיים וחוויות אסתטיות. החלק הראשון מתמודד עם האפשרות של קבלה רציונלית של עמדת הזולת בעניינים אסתטיים, ואילו השני עוסק באפשרותה של הערכה אסתטית שאינה מבוססת על תפיסה חושית. השער הרביעי והאחרון של חיבור זה דן באי הרלבנטיות של חלופות בלתי רלבנטיות בסוגיות של בחירה. הדיון מציע סטנדרט חדש ומרחב לבחירות רציונליות אותו אני מכנה תועלת בחירה. במהלך הדיון אציג ביקורת כלפי הסטנדרט המקובל של תוצאתנות וקונסיסטנטיות פנימית.

ביסודו של חיבור זה מונח הרעיון של רציונליות מוגבלת או לקויה. בניגוד לאופן החשיבה ה"אידיאלי", שבדיון על אודותיו לרוב אין דנים בשיקולים קוגניטיביים, טמפורליים ומעשיים ובנקודות מבט שונות שמתוכן חשיבה נעשית, הרי שהמאפיינים הללו, המבדילים את אופן החשיבה האנושי מזה האידיאלי, מצויים במרכזו של חיבור זה. בהתאמה, החיבור יציג מודל רציונליות שעל בסיסו ניתן לבחון ולנתח את החשיבה האנושית. במידה רבה מטרתו של פרויקט זה היא לתחם באופן מדויק מתי, כיצד ואיך שני סוגי הרציונליות הללו – האנושי מזה והאידיאלי מזה – נבדלים זה מזה, ומתי והיכן חל כל אחד מהם בעת בחינה ביקורתית של האמנות ושל התנהגות.

הבעיות הנדונות בחיבור עוסקות, בעיקרן, בשאלה כיצד עלינו להתייחס למידע ולראיות לאור הידיעה שלעתים קרובות אנו מעניקים להם פרשנות שגויה. במילים אחרות, בפרויקט זה אני מבקש לפתח מודל חדש של אפיסטמולוגיה פרסקריפטיבית. כמו אתיקה, זוהי דיסציפלינה שמספקת לנו הצעות מעשיות לחיים טובים יותר. המודל האפיסטמולוגי שאני מבקש לפתח אינו מספק כללי התנהגות או כללים לניהול עצמי, אך הוא מצייד אותנו בעקרונות שבכוחם לסייע לנו בבחירותינו ובאופן בו אנו מקבלים החלטות באמצעות זיהוי מגמות שכיחות של דפוסי חשיבה שגויים. ישנם שני טעמים מרכזיים לחשיבותו הרבה של המודל. ראשית, החשיבה שהמודל עוסק בדפוסיה מלווה אותנו כל העת, ולפיכך יש ערך רב לידיעה כיצד לחשוב נכון. שנית, ואולי חשוב יותר, חשיבה שגויה הרבה יותר מסוכנת ומזיקה ממה שנדמה לנו. בהתאם, חילוצם של דפוסי השגיאות המתקבלים כתוצאה מחשיבה שגויה תסייע בהדגשת הצורך להימנע מחשיבה לקויה.

בחיבור זה שישה פרקים. הפרק הראשון, "הקושי בפתרונות הומוגניים לאי הסכמה בין בני פלוגתא שווי מעמד ויכולות", מבקש לערער על הגישה הרווחת בספרות הפילוסופית, לפיה יש לבקש אחר פתרון הומוגני לכל מקרה של אי הסכמה בין בני פלוגתא שווי מעמד ויכולות. טענתי היא שאי הסכמות בשדות שונים צריכות להיפתר באופנים שונים. כלומר, כאשר אנו נוכחים לדעת שאדם שמיומנותו בתחום הרלבנטי שקולה לשלנו, מחזיק – על בסיס אותן הראיות – בעמדה שונה משלנו בנושא מסוים, אזי צורת ההתייחסות שלנו למקרה צריכה להיות שונה בהתאם לתחום. את ההצעה הקונסטרוקטיבית שאני מציג בפרק ניתן לתמצת באופן הבא: במקרים מסוג זה, הסיכוי שאדם ישנה את דעתו אל מול אי הסכמה מצד עמיתו גדל ככל שההליכים לבחינת הראיות גלויים יותר וככל שהראיות בתחום מסוים פחות עמומות. מסקנתו של הפרק היא ששדות שונים מאפשרים פשרות אפיסטמיות מסוגים שונים.

הפרק השני, "הקושי האמיתי בסוגיה של אי הסכמה", מציג ערוץ חדש להבנתן של הסוגיות הבסיסיות הכרוכות בדיון בדבר משמעותה האפיסטמית של אי הסכמה. הפרק מבקש לחלץ את הבעייתיות האמיתית הכרוכה באי הסכמה. ניתן לתאר בעייתיות זו כקושי להחליט האם בחינתן מחדש של ההאמנות שברשותי, בעקבות גילוייה של אי הסכמה, היא טקטיקה נכונה; וכקושי להחליט באיזו טקטיקה של בחינה מחדשת עלי לנקוט. הפרק מראה כי מדובר בקושי בלתי נמנע, שקשה מאוד "לשים עליו את האצבע", ושאי אפשר לדון על אודות

אי הסכמה אינו יכול להתחמק ממנו. הפרק מתקדם בשני שלבים: בשלב הראשון אציג את הבעייתיות האמיתית בסוגיה של אי הסכמה. בשלב השני אבחן את שלוש הגישות הדומיננטיות בספרות הפילוסופית לאי הסכמה, וכן את יתרונותיה של בחינת העמדות המחודשת שכל אחת מהן מציעה. הדיון יראה כי הגישות הללו ואופן הפעולה שהן מבקשות לקדם אינם עוסקים למעשה בבעייתיות האמיתית בהקשר של אי הסכמה. במהלך הדיון אצביע גם על הגישה הסבירה ביותר מבין השלוש על בסיס הנימוק שהיא מזהה, במובלע, את הבעייתיות המוצגת בחלקו הראשון של הפרק.

הפרק השלישי, "ערך אפיסטמי ופשרה אפיסטמית, תשובה לשרה מוס," דן בהצעה שניתן להגיע לפשרה אפיסטמית תוך התייחסות לערך האפיסטמי שניתן לערך הסתברותי ולא רק לערך ההסתברותי עצמו. מוס מציעה להשתמש בחוקי ציון לפיהם כל צד במחלוקת יצינן את הערך האפיסטמי של כל הערכה הסתברותית שהוא נותן כך שפשרה אפיסטמית בין הצדדים תהא כאשר הערכים האפיסטמיים של ההסתברויות כמו גם ההסתברויות עצמן נלקחות בחשבון. בפרק זה אציג שני טיעונים מדוע ההצעה של מוס בעייתית.

הפרק הרביעי, "העדפת דעתו של הזולת בעניינים אסתטיים", עוסק בשאלה האם יש מקרים בהם העדפת דעתו של הזולת בעניינים אסתטיים עשויה להיחשב רציונלית, והאם יש הצדקה אפיסטמית – בהקשר של שיפוטיות אסתטיים – להמעטה של אדם בדעתו שלו, או לביטולה, בגין עמדתו של הזולת באותו עניין. טענתי היא שהעדפת דעתו של הזולת בעניינים אסתטיים יכולה, בתנאים מסוימים, להיחשב רציונלית. במהלך הדיון הפרק שב ועוסק בשתי קונספציות רווחות, בעלות זיקה הדדית. הקונספציה הראשונה כוללת טענות מוכרות ביחס למוצאן הסובייקטיבי של תפיסות אסתטיות, ולפיה שיפוטיות אסתטיים של הזולת לרוב אינם נתפסים כבסיס ליצירתן של תפיסות אסתטיות. יתרה מכך, כאשר שיפוטיות של הזולת יוצרים בסיס שכזה, הרי שלרוב – בהקשר בו אנו נדון – הדבר נחשב, ולו במובן מסוים, לבלתי לגיטימי. לפי הקונספציה השנייה, שאלות אפיסטמולוגיות על אודות שיפוטיות אסתטיים גוררות עוגנים מטפיזיים בדמותן של מחויבויות מטפיזיות בדבר ריאליזם אסתטי. אין זה קוהרנטי אפוא, לפי קונספציה זו, לטעון שיש להעדיף שיפוט אסתטי מסוים מבלי להודות בממשותן של תכונות אסתטיות. בהתאמה, הטענה ששיפוט אסתטי מסוים הוא טוב יותר, ושלכן יש להעדיפו על פני אחר, גוררת את ההנחה שקיים סטנדרט שיפוטי שעל בסיסו אפשר לטעון טענה שכזו. קיומו

של סטנדרט שכזה מניח נקודת משען בלתי תלויה, תמיכה שרק אמצעי מטפיזי מסוגל להעניקה. בעייתיותה של הנחה זו נדונה כבר במסגרת הגישה האנטי ריאליסטית, הגישה הדומיננטית בספרות, בהיבט של תכונות אסתטיות.

כאמור, הפרק יקיים דיון מחודש בשתי הטענות המוכרות הללו, אך יחד עם זאת, טענתי כי בהינתן תנאים מסוימים העדפת דעתו של הזולת בעניינים אסתטיים יכולה להיחשב רציונלית, איננה טענה חדשה. העדפת עמדתו של הזולת בעניינים אסתטיים היא פרקטיקה שגרתית למדי. חדשנותו של הטיעון בפרק זה מתבטאת בכך שהוא מראה כיצד העדפת עמדתו של הזולת בעניינים אסתטיים היא מן האפשרי; כלומר, זהו ניסיון להצדיק מקרים בהם אנו מעדיפים לאמץ את דעותיו של הזולת בעניינים אסתטיים. לסיכום, לפרק זה שתי מטרות: האחת, להראות שבמקרים מסוימים העדפת עמדתו של הזולת בהקשר בו אנו דנים היא פרקטיקה רציונלית; והשנייה, להציע מערכת תנאים שיאפיינו את המקרים הללו כפרקטיקה רציונלית.

הפרק החמישי, "עיקרון ההיכרות, אוטונומיה אסתטית והערכה אסתטית", מבקש להסביר את אפשרותה של מה שאכנה הערכה אסתטית שאינה מבוססת-תפיסה חושית, וכך לבסס את הרציונליות המונחת ביסודן של חוויות אסתטיות מיצירות אמנות קונספטואליות, שהערכתן אינה מעוגנת בתפיסה חושית. עיקרון ההיכרות, שלפיו לא ייתכן שתפיסותיו האסתטיות של מאן שהוא יתבססו על חווייתו האסתטית של הזולת, נקשר לאחרונה לדיון על אפשרותה של תזת העדות בדבר האסתטי. במילים פשוטות, תזה זו עוסקת בשאלה האם ניתן לרכוש ידע בדבר האסתטי באמצעות עדות, עדותו של אדם אחד בפני אחר – לרוב תיאורים על האסתטי ותיאורים שאינם מצויים בזיקה לאסתטי. בהקשר זה הועלו כמה וכמה הצעות שתמציתן קבלה מצומצמת של עיקרון ההיכרות. פרק זה מבקש להראות שיש לצמצם את העיקרון אף יותר. בפרק זה שני חלקים: בחלק הראשון אסביר את העיקרון, אדון ביסודותיו המושגיים ואציג מספר פרשנויות מהספרות העכשווית לעיקרון ולעמדה שהוא מבטא. בחלק השני אטען, באמצעות מתן דוגמאות, שהעיקרון ניתן ליישומים מגוונים בתחומים שונים.

הפרק השישי, "ערכי תוצאה וערכי החלטה - האי רלבנטיות של חלופות בלתי רלבנטיות בסוגיות של בחירה",
עוסק בסטנדרטים של בחירות רציונליות. הפרק מציע הבחנה בין ערך התוצאתי של בחירה וערך ההחלטה
בפני עצמה במצבי בחירה. בעשותו כן, הפרק מציע טסקונומיה חדשה להתמודדות עם מצבי בחירה. מודגם
כיצד ערך החלטה יכול להסביר מגוון רחב של בחירות שכאשר הן נבחנות על פי ערכן התוצאתי הן מתפרשות
כאירציונליות. הפרק מציע כלים חדשים הערכה של בחירות רציונליות בהיותו טוען שהסטנדרט המקובל הוא
חזק מדי.

מחלוקות רציונליות ופשרות אפיסטמיות:
עקרונות של שיפוט וקבלת החלטות בתנאי אי-ודאות

חיבות לשם קבלת תואר דוקטור לפילוסופיה

מאת אמיר קוניגסברג

הוגש לסנט האוניברסיטה העברית בירושלים בחודש מרץ 2012

מחלוקות רציונליות ופשרות אפיסטמיות:
עקרונות של שיפוט וקבלת החלטות בתנאי אי-ודאות

חיבות לשם קבלת תואר דוקטור לפילוסופיה

מאת אמיר קוניגסברג

הוגש לסנט האוניברסיטה העברית בירושלים בחודש מרץ 2012

עבודה זו נעשתה בהדרכתם של איל וינטר, משה הלברטל, וגבריאל מוצקין