‘On Certainty’ and Formal Epistemology

MSc Thesis *(Afstudeerscriptie)*

written by

Moritz M. Bäumel
(born 14.09.1986 in Munich)

under the supervision of **Dr. Sonja Smets and Prof. Dr. Martin Stokhof**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam*.

Date of the public defense:	**Members of the Thesis Committee:**
June 23rd, 2015	Dr. Maria Aloni (Chair)
	Dr. Sonja Smets (Supervisor)
	Prof. Dr. Martin Stokhof (Supervisor)
	Prof. Dr. Franz Berto
	Prof. Dr. Michiel van Lambalgen
	Dr. Bryan Renne

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

We look at Wittgenstein's epistemological work *On Certainty* from the point of view of formal epistemology. Different interpretations of the work are looked at and important aspects of them are worked out. We take probability theory as a basis and interpret probabilities in terms of betting quotients, i.e. as a basis for action, and define certitude as probability 1. On top of this we define such notions as belief, doubt and knowledge in the spirit of *On Certainty*, following work done by A. Baltag. We proceed by looking at ways of (i) extending this system to a multi-agent system, (ii) outfitting it with an internal perspective and (iii) dynamics. We also look at game-theoretic approaches. The merits of the work will be discussed, as well as shortcomings and future work to be done.

Acknowledgements

I would first like to thank my logic teachers who have sent me on this journey: (in order of appearance) Dr. Behboud, Dr. Centrone, Dr. Krämer, Prof. Veltman, Prof. van Rooij, and Dr. Baltag. Moreover, of course, thanks goes to my supervisors Dr. Sonja Smets and Prof. Martin Stokhof for encouraging me to take on this project and having time to meet me in a time that, for both of them, was very busy and hectic. I am also indebted to Prof. Löwe and the staff at the ILLC for help and hints whenever I needed it; a special thanks also goes to Tanja Kassenaar. With respect to this theses, Prof. van Lambalgen's work on Kant has been a great source of inspiration and Dr. Baltag's work was integral, so thank you for that. In general, I have to thank the ILLC for creating an environment in which original thought and research is nurtured and encouraged, rather than requiring students to regurgitate the research of others; this is a very unique place.

On a personal note, I would like to thank all my fellow logic students at the ILLC, especially Konastantinos Gickas and Aldo Ramírez Abraca for the great team projects we did together — also at this point thanks to Paolo Galeazzi for the wonderful supervision (you'll make a great teacher!). With the great time consumption this program has demanded and the little free time thus at hand, I want to especially thank my Leonie for sticking this one out with me; I know it wasn't fun for you. Last but not least I would like to thank the greatest scholarship provider in the world for the support (not just financially) and encouragement throughout my studies — thanks, mum!

Contents

Abbreviations

CPR I. KANT [1998]. *Critique of Pure Reason*. Ed. and trans. by P. Guyer and A. Wood. Cambridge University Press, UK.

OC L. WITTGENSTEIN [1991]. *On Certainty*. Ed. by G. Anscombe and G. von Wright. A Blackwell Paperback. Wiley.

PG L. WITTGENSTEIN [1974]. *Philosophical Grammar*. Ed. by R. Rhees. Basil Blackwell, Oxford.

PI L. WITTGENSTEIN [2010]. *Philosophical Investigations*. Ed. by P. Hacker and J. Schulte. Wiley.

Introduction

On Certainty (henceforth referred to as *OC*) is Wittgenstein's last work. Written as a series of notebook entries, the last entry was written only two days before Wittgenstein died on April 29^{th}, 1951 [cf. OC, 670–676]. Primarily concerned with issues of epistemology, *OC* is Wittgenstein's response to the problem of the sceptic, i.e. can we *really* know anything? Wittgenstein's response, roughly, is that there are things we can be certain (called certainties or hinges) of and that we are certain of them is simply apparent in our lives, the way we act, etc.

It is questionable whether *OC* produces a full theory of knowledge (it can even be questioned whether it produces any theory). Nonetheless, it is an interesting work in epistemology for the observations it makes. It is undoubtedly correct to say that Wittgenstein is vastly studied by philosophers (among others), especially his philosophy of language. His thoughts on epistemology as collected in *OC* are now increasingly being studied as well. Moyal-Sharrock and Brenner [2007, p. 2] collect a series of articles that discuss different interpretations *OC* gives rise to:

> The increasing interest in *On Certainty* has manifested itself in efforts at understanding (1) the work itself; (2) its place in Wittgensteins philosophy; (c) [*sic.*] its relevance for philosophy in general, and epistemology in particular.

In this thesis we are concerned with the first and, mainly, the third effort. In particular, we aim at generating interest in studying *OC* from the point of view of formal epistemology. Formal epistemology, roughly explained, is concerned with using mathematical or logical methods to formulate theories of epistemic notions such as knowledge, belief, etc. The methods used include, among others, the epistemic (modal) logic, dynamic epistemic logic, probability theory and game theory. As we shall see shortly, we will make use of all these methods in order to interpret *OC* formally. This is a daring attempt, and we do not claim (or aim) to present the reader here with a sound and complete logic of *OC*. Rather, we provide tools and techniques that quite naturally fit with Wittgenstein's aphorisms. The hope is to inspire further research; both further formal treatments of *OC* and further interpretations of formal methods in a Wittgensteinian manner. Criticism of what is to follow is thus to be expected, and welcome.

This is how we proceed. In chapter 1 we will look at *OC* itself and begin by introducing the sceptical problem, G.E. Moore's response to it and Wittgenstein's response to Moore. In section 1 we follow Moyal-Sharrock and Brenner [2007] and lay out the different interpretations *OC* has given rise to. We begin in subsection 1.1 with a foundationalist reading, which

will be subjected to criticism. We conclude there that Wittgenstein was not a foundationalist, at least not in the traditional sense of the word. In subsection 1.2 we look at a Kantian, transcendental reading, which we conclude to be largely concerned with metaphysical issues and, thus — although interesting issues are raised —, not as relevant to our current project. Following this, we look at a therapeutic reading in subsection 1.3 which questions whether Wittgenstein makes any theoretical claims at all. We conclude that Wittgenstein is, at least in part, aiming for a theory. Finally, we look at the so-called epistemic reading in subsection 1.4 which is concerned with the epistemic status of certain notions important to *OC*. In section 2 we collect certain aspects that turn out to be important throughout section 1. We begin in subsection 2.1 to look at the very important notion of a hinge proposition, or certainty; in subsection 2.2 we look at cases where these two people's hinges are incompatible. In subsection 2.3 we look at the distinction Wittgenstein makes between first- and third-person knowledge claims and, finally, in subsection 2.4 we look more closely at the very important notions of knowledge, belief, certainty and doubt and how they might interact.

Having worked out the aspects that to us seem most important, we turn to giving them a formal treatment in chapter 2. We simply take certitude to be probability 1, and argue for this by interpreting probabilities as a basis for action. We will stay entirely on a semantic level throughout this thesis. In section 3 we explore a static probabilistic epistemic logic that is due to Baltag [unpublished] which, when probabilities are interpreted in terms of betting quotients (as we will do), provides a good basis to formalize the (epistemic) notions important to Wittgenstein. We begin by looking at certainty in subsection 3.1, then turn to belief and doubt in subsection 3.2 and in subsection 3.3 we look at knowledge which requires an objective verification, as opposed to certainty, belief and doubt. In section 4 we provide tools that are capable of modelling the other aspects we have worked out such as disagreement on the level of certainties, a first-person perspective (as opposed to the third-person perspective of section 3), and a way to handle knowledge-claims. We will also provide suggestions as to how we could integrate such approaches into the setting of section 3. We begin by looking at ways to extent the setting to include multiple agents in subsection 4.1. Next, we introduce a first-person perspective following Aucher [2008] and suggest a way to adapt this to a probabilistic setting in subsection 4.2. In the last section of the chapter we look at ways to handle knowledge claims. We discuss various approaches to handle so-called updates that change the information state of an agent but also look at a game-theoretical approach.

After two long chapters it is time to reflect. We will do so in chapter 3 where we first ask in section 5 what the merit of such a formal approach to *OC* is. We conclude that this formal approach can not only clarify *OC*, but also make us look at formal techniques from a Wittgensteinian point of view. In section 6 we will deal with a problem that will pop up throughout chapter 1, *viz.* in how far is logic certain or known? It turns out that Wittgenstein does not seem to take logic as fast-standing as our system does and that including this is a problem that our setting cannot properly deal

with. In section 7 we will mention some more ways to get a formal hold of even more aspects of *OC*, before we finally conclude.

CHAPTER 1

Aspects and Interpretations of *On Certainty*

1. Interpretations

INTRODUCTION. *OC* starts out as a reply to G.E. Moore's *Proof of an External World*. There, Moore [1939] offers a solution to the problem of the sceptic. Roughly, the sceptical problem arises from the following, intuitively acceptable but incompatible assumptions [cf. Pritchard 2007, p. 189]:

(1) We are unable to know the denials of the sceptical hypothesis [that there are no physical objects].
(2) If we are unable to know the denials of sceptical hypotheses, then we are also unable to know any one of the 'everyday' propositions which we typically take ourselves to know.
(3) We are able to know everyday propositions.

These claims are clearly incompatible as the negation of (3) follows from (1) and (2) by *modus ponens*. The sceptic will deny (3) as she accepts that the sceptical hypothesis cannot be known to be false. In a nutshell, Moore argues the contrary by applying a logical move sometimes called a *Moore Shift*.[1] Arguing from common sense, he deduces that (3) is correct and that, accepting (2), by *modus tollens* it follows that we can, and in fact do, know the denial of the sceptical hypothesis.

Surely, this argument rests on the fact that Moore *knows* some everyday propositions. His (in)famous example of such a known, everyday proposition is, while holding up ones hand, uttering 'Here is a hand.' This claim is the starting point for *OC*. Wittgenstein writes:

> If you do know that *here is one hand*, we'll grant you all the rest. [OC, 1]

Here, Wittgenstein seems to accept (2) above as well as the Moore Shift as a form of argument as this quote is just the contrapositive of (2). What Wittgenstein takes issue with is Moore's claim to *know* that 'Here is a hand.'

> From its *seeming* to me or to everyone to be so, it doesn't follow that it *is* so.
>
> What we can ask is whether it can make sense to doubt it. [OC, 2]

Wittgenstein agrees with Moore in so far as he does not doubt that there is a hand, but he does not accept Moore actually knows this in a way relevant for his argument to go through. Wittgenstein accepts that any proposition that is truly known must be true but that we cannot subjectively establish that we know something in the way that Moore does.

[1] To my knowledge, this term was coined in Rowe and Trakakis [2007, p. 65].

> —For 'I know' seems to describe a state of affairs which guarantees what is known, guarantees it as a fact. One always forgets the expression 'I thought I knew'. [OC, 12]

Knowledge-claims are defeasible for Wittgenstein; one can be under the impression to know something without actually knowing it. Therefore, their utterance that they know a proposition cannot establish the truth of that proposition, for they could be mistaken.

> For it is not as though the proposition 'It is so' could be inferred from someone else's utterance: 'I know it is so'. Nor from the utterance together with its not being a lie.—But can't I infer 'It is so' from my own utterance 'I know etc.'? Yes; and also 'There is a hand there' follows from the proposition 'He knows that there's a hand there'. But from his utterance 'I know. . .' it does not follow that he does know it. [OC, 13]
>
> That he does know remains to be shewn. [OC, 14]

However, Wittgenstein makes it clear that he does not doubt that there is a hand [cf. OC, 2].

> It's not a matter of *Moore's* knowing that there's a hand there, but rather we should not understand him if he were to say 'Of course I may be wrong about this'. We should ask 'What is it like to make such a mistake as that?'—e.g. what's it like to discover that it was a mistake? [OC, 32]

According to Wittgenstein, Moore does not *know* that there is a hand, but he is also not mistaken in uttering 'Here is a hand.' How could this be?

> It maybe for example that *all enquiry on our part* is set so as to exempt certain propositions from doubt, if they are ever formulated. They lie apart from the route travelled by enquiry. [OC, 88]

1.1. THE FRAMEWORK READING. The quote from OC, 88 may suggest that there are certain propositions that have a special place among all propositions in that they cannot be doubted, they form a *foundation* for all our enquiry. In how far can we, then, read *OC* as a foundationalist work?

Stroll [1994, pp. 144–8] metaphorically likens the view of (epistemic) foundationalism to the picture of an inverted pyramid. The bottom, comprising a small class of propositions (or judgements) is the foundation of the larger structure on top of it, every proposition (or judgement) in the top is justified by a lower one. The very bottom of the pyramid comprises the class of *basic* beliefs which, to avoid infinite regress and thereby open the door to scepticism, are taken to be *intrinsically credible*. This intrinsic credibility traditionally sharply restricts the class of basic beliefs in terms of content and thus it becomes a problem for the foundationalist to build the superstructure (the whole pyramid) on the basis of this content-wise limited class of beliefs. For example, we could argue, with a phenomenalistic foundationalist, that knowledge-claims about how things appear to us are intrinsically credible as we are unable to make clear sense of what it would mean for things to only *appear to appear* to us in a certain way. Granted this

intrinsic credibility of experiential knowledge and given an inferential connection between our foundational knowledge and the non-foundational knowledge, however, it is still unclear how experiential knowledge can be used to ground our knowledge of everyday propositions about external things as, say, 'water boils at 100° C' [cf. Williams 2007, p. 48].

Wittgenstein avoids limiting his certainties in terms of content.

> —As if giving grounds did not come to an end sometime. But the end is not an ungrounded presupposition: it is an ungrounded way of acting. [OC, 110]

With his concept of meaning as use, Wittgenstein thus offers an account of foundations that is far less limited in content.

> Framework judgements are not certain because self-evident or intrinsically credible. Rather, their certainty accrues to them as a matter of meaning. [Williams 2007, p. 49]

Perhaps, then, we can take these certainties as a foundation in the foundationalist sense and not run into the problem described above. To get back to our (very simplistic) example above, my knowledge that water boils at 100° C may not need to find grounding in an experiential proposition, but rather shows itself in my boiling water every day to cook. Whether this can count as a foundationalist theory, however, needs to be shown.

From Stroll's simple picture, Williams draws some far-reaching conclusions: (i) "the image definitely suggests the foundations of knowledge are the same for everyone"; (ii) "the image invites us to suppose that we can draw a line across the pyramid, separating the base from the superstructure"; (iii) "the image makes it natural to suppose that, with the line in place, we could cut off the top of the pyramid, leaving its more 'basic' section intact"; (iv) "the picture of the superstructure resting squarely on the base, hints strongly at some kind of tight logical connection between basic and non-basic beliefs" [*ibid.*, pp. 50–1]. To be fair to the foundationalist here, Williams seems to get ahead of himself in point (i) suggesting that foundations need to be universal. Williams [*ibid.*, p. 50] argues this universality is needed to make foundationalism "a theory of *knowledge*." The emphasis on 'knowledge' here, I take it, already precludes a relativist view. From a structural perspective, however, there is nothing intrinsic to foundationalism ruling out a relativist stance, as Williams [*ibid.*, pp. 55–6] himself admits, although it would be fairly untraditional.[2]

Williams [2007, p. 51] calls the "four characteristics of traditional foundationalism" corresponding to points (i)–(iv) above, respectively, "universality, specifiability, independence and adequacy". He argues that Wittgenstein would agree with none of them and elsewhere even Stroll [2007, p. 34] agrees that "[t]he idea that Wittgenstein is a foundationalist is highly controversial, since in the *Investigations* he rejects any form of foundationalism."

[2] On this topic see also Krausz [1984], who distinguishes relativism (a stance on truth and, thus, logic) from foundationalism (an epistemological theory) and discusses connections of them and their negations.

We begin, with Williams, by looking at specifiability. He writes that "Wittgenstein denies that the class of basic certainties can be theoretically delimited" [2007, p. 51], which is a necessary condition for the traditional foundationalist to form a "*theory* of knowledge" [*ibid.*, p. 50]. That is to say, with no theoretical way[3] of delimiting intrinsic credibility from inferential credibility, foundationalism is missing something. Evidence of this denial can arguably be found in *OC*, where Wittgenstein writes:

> One may be wrong even about 'there being a hand here'. Only in particular circumstances is it impossible.— 'Even in a calculation one can be wrong — only in certain circumstances one can't.' [OC, 25]
> But can it be seen from a *rule* what circumstances logically exclude a mistake in the employment of rules of calculation?
> What use is a rule to us here? Mightn't we (in turn) go wrong in applying it? [OC, 26]

We can see that Wittgenstein in OC, 25 takes it that a proposition's certainty-status depends on surrounding circumstances (at least in some cases);[4] i.e. a proposition can be certain in some circumstance and uncertain in another. In considering the possibility of a rule in OC, 26 that is to tell us under which circumstances some particular proposition is certain, Williams sees a return to the *Philosophical Investigations* (henceforth *PI*). A central theme there is that "language cannot be guided by explicitly formulated rules, for such rules would themselves be open to misinterpretation or misuse" [Williams 2007, p. 52]. With respect to *OC* this means that no rule can *with certainty* establish whether a proposition is certain in a given circumstance (or at least not with a certainty greater than the proposition). Wittgenstein makes it clear that we cannot describe such rules without reference to "normal circumstances" and that we can, "[a]t most, [...] describe a range of abnormal ones" [OC, 27]. Williams [2007, p. 52] concludes:

> It is not just that in practice the boundary between a certainty and a hypothesis is vague — though this is also true — it is that, at a general theoretical level, it is unspecifiable in principle, even vaguely.

Thus, if Wittgenstein were a foundationalist, he would not be producing a foundationalist *theory* of knowledge in the traditional sense, as he has no way of theoretically demarcating certainties from other propositions. This, however, is important to (at least absolutist) foundationalists, as universal foundations would require "some way [...] of delimiting the class of basic beliefs" according to Williams [*ibid.*, p. 51]. This is not a very strong point. The relativist immediately has a way out and even the absolutist might not be convinced. Williams seems to base the requirement of a theoretical delimitation of basic and non-basic beliefs on the fact that "all traditional foundationalisms propose some such principle of demarcation" [*ibid.*, p. 50]. To be fair to Williams, however, he is not (yet) claiming to have

[3] 'Theoretical', here, is understood very broadly as the *possibility* to "provide some way, however vague, of delimiting the class of basic beliefs" [*ibid.*, p. 51].

[4] See sections 1.4 and 2.1.

shown Wittgenstein is not a foundationalist; he merely aims to show that Wittgenstein is no *traditional* foundationalist.

Stroll [1994, p. 155 ff.] then argues in defence of a foundationalist reading that we must distinguish two types of foundationalism found in *OC*. First, we have what Stroll calls the *propositional* foundationalism which is more akin to traditional foundationalism in that it identifies propositions (or judgements) at the bottom of our pool of knowledge. Second, there is a *non-propositional* foundationalism, more prominent later in *OC*, that takes the basis of knowledge to be found in action — not propositions or judgements — and that it is in this way that we can demarcate the foundations from the superstructure: the foundations are ways of acting, the superstructure claims of knowledge based on them. This second view is supposedly to be found in the following passage:

> Giving grounds, however, justifying the evidence, comes to an end;—but the end is not certain propositions' striking us immediately as true, i.e. it is not a kind of *seeing* on our part; it is our *acting*, which lies at the bottom of the language-game. [OC, 204]

However, as Williams [2007, p. 53] points out, the contrast in the above quote is "not between acting and *judging* but between acting and *seeing*" as Wittgenstein's own emphasis all but confirms.

> Certainties are not held fast because of their transparent truth (to the eye of the mind) but because of their role in language-games. They are certain because they are treated as such. [. . .] Judgements would not mean what they do [. . .] if [certainties] were treated as open to doubt. [*ibid.*]

The point here, as far as I can see, is that in this case, too, there is no demarcation between acting and judging (judging *is* acting), and thus no delimitation between the foundation (actions) and the superstructure of (grounded) knowledge-claims (judgements). There is overlap between the foundations and the superstructure, but that is not to say that there is no way to distinguish knowledge from certainty: certainties "are certain because they are treated as such" [*ibid.*] but knowledge requires *truth*. The clear cut demarcation of traditional foundationalism is, however, not to be found here either.

None of what Williams writes so far has invalidated the picture of the inverted pyramid for *OC*. That is to say, the foundationalist interpretation still holds, albeit a fairly untraditional form of foundationalism. The certainties are still at the bottom of the pyramid with the remainder of our beliefs resting on top. And while we cannot delimit them (with a rule) from other propositions, Wittgenstein does give us a wealth of examples. These examples — e.g. 'here is a hand', 'the earth has existed long before I was born' —, however, do not form a homogeneous whole in terms of content.[5] Moreover, Wittgenstein seems to include propositions referring to history,

[5]Certainties *do* perform at least a similar function in a language-game but can function differently in different contexts.

physics, etc.[6] and these seem to be able to be culture-specific. Thus, it seems hard to maintain that *all* basic certainties are universal (though some certainly are).[7]

Moving on, Williams [*ibid.*, p. 53] argues that Wittgenstein's identification of judgement as an action "links his approach to certainty with his conception of meaning-as-use" which, in turn, "implies a limited semantic holism." This suggests that independence of certainties from other, non-certain propositions which is vital to the foundationalist project[8] is not present in *OC*. And indeed, as the next passages seems to suggest, Wittgenstein does not take certainties as semantically or epistemically independent, at least from one another:

> We do not learn the practice of making empirical judgements by learning rules: we are taught *judgments* and their connection with other judgments. A *totality* of judgments is made plausible to us. [OC, 140]
> When we first begin to *believe* anything, what we believe is not a single proposition but a whole system of propositions. (Light dawns gradually over the whole.) [OC, 141]

"Practices must be mastered whole: basic certainties are held in place by things around them" argues Williams [2007, p. 54]. Williams [*ibid.*, p. 55] sees expressed in the above quote an "essential semantic interdependence of certainties and non-certainties." I would argue that this is not the case. The holism here, it should be stressed, is *limited*. It should not invoke the impression that certainties are "held in place" by ordinary, non-certain propositions; rather, they form a tightly knit web among themselves.[9] Wittgenstein [OC, 152] likens certainties to the axis around which a body may rotate and stresses that the "axis is not fixed in the sense that anything holds it fast, but the movement around it determines its immobility." The web of certainties is not held in place by non-certain propositions but its status as a web of *certainties* (as opposed to just any web of interdependent propositions) is dependent on non-certain propositions floating around it. The web, as tightly knit as it is, however, is dynamic in that it allows for new propositions to enter and for (some) certainties to detach [compare also the metaphor of the riverbed OC, 95–99]. It is this dynamic that lets me agree with Williams [2007, p. 55] that puts Wittgenstein at odds with foundationalists, as he denies a status of "non-inferential entitlement to particular propositions" they traditionally require and thereby denies certainties to be independent from non-certainties.[10] Some certainties can, at times and

[6] Cf., e.g., OC, 183: "'It is certain that after the battle of Austerlitz Napoleon... Well, in that case it's surely also certain that the earth existed then."'

[7] Cf. Williams [2007, pp. 53–54] and sections 1.4 and 2.1.

[8] It is vital as "there would be little point in insisting that the superstructure rests on the basis if, in order to entertain 'basic' certainties, we had to presuppose lots of other non-basic propositions to be true" [*ibid.*, p. 50]

[9] This will be further explored in subsection 2.1.

[10] On a somewhat related note, Wittgenstein is also at odds here with the coherence theorist — the standard denial of foundationalism. Coherence theory "recognizing semantic inter-dependence, concludes that no epistemic entitlements are genuinely non-inferential" [*ibid.*]. Wittgenstein, on the other hand, makes it clear that there is a hard rock bottom of the

in unusual circumstances, demand grounding (possibly even from non-certain propositions). For example, 'My name is M.B.' would count as certain for me as I prove day-in and day-out in my acting. In the unlikely (though possible) event that someone tells my I only have amnesia, I might pull out my ID card and thereby produce a justification. Whether I can rely on my ID with a certainty greater than my knowing my own name is at least disputable. Independence, thus, is on shaky ground in *OC*.

In an effort to argue rational adequacy is not found in *OC*, Williams [2007, 55ff.] refers us to one of Wittgenstein's own examples [cf. OC, 92]: Moore meets a king who was made to believe the world to have begun with his birth. How is Moore to convince the king of the contrary certainty that the world was long there before the king? The point is that while Moore and the king will undeniably share some common ground, they clearly do not share the certainty that the world has existed for a long time before their birth. In order to convince the king of this, Moore would need to make this (supposed) fact certain to the king. But then "the king would be brought to look at the world in a different way" [OC, 92]. That is, if the certainties form the basis of our language-game, if they are the background against which true and false is judged [cf. OC, 94, 204, 403], there is no way for Moore to convince the king by *rational argument*; "The system [of certainties] is not so much the point of departure, as the element in which arguments have their life" [OC, 105]. Thus, rational adequacy, as required by the foundationalist, has no place in *OC* in this form. The superstructure of beliefs does not, by means of rational adequacy, rest on the foundations. This is quite the blow, as, as Williams [2007, p. 51] argues, there has to be a *logical* connection between the superstructure and its foundation, for were it not logical, then the justificatory force of the foundations would rest on something that might be in need of justification itself. Take our example from before. If we accept experiential claims as basic, we need some principle, say, 'however I perceive things is how they are' in order to gain empirical knowledge on the basis of experiential knowledge. The principle itself is not experiential and hence not intrinsically credible, thus the foundations loose their justificatory force.

Now, certainties, it seems, are the foundation of reasoning itself, they lay out the logic of our language. The foundationalist might reply 'eureka!' Wittgenstein has given him a way out of a problem once again. He can drop the requirement that the connection between superstructure and foundations be purely logical and merely require the connection to be found *within* the foundation.[11] After all, this is what Wittgenstein seemingly does by saying certainties are "part of judging" and "part of our *method*" [OC, 150–1], that is, getting from the foundation (our actions or certainties) to fully

---

riverbed, an immovable core of the web that is part of the logic of our language and cannot change or be questioned under any circumstance. At least not under any circumstance brought about *by* the logic of our language; outside influences may be able to change this. Sticking with the riverbed-metaphor, we can think of an earthquake shaking up even the most hardened certainties.

[11]This would, in effect, show what a combination of foundationalism and relativism might amount to.

formed judgements is achieved by means that are just as certain. Wittgenstein, when discussing the example above, writes:

> Remember that one is sometimes convinced of the correctness of a view by its *simplicity* or *symmetry*, i.e., these are what induce one to go over to this point of view. [OC, 92]

The foundationalist might now maintain that the connection is to be found in the certainties themselves and that things like a longing for symmetry and simplicity are evident in our action. Nevertheless, if Moore can convince the king of revising one certainty, he might need to revise the whole web of certainties he has — "the king would be brought to look at the world in a different way." The web of certainties resulting from the king's old web together with the insight that the world has existed before his birth (which he accepts on the basis of, say, simplicity), may not yet be in line with Moore's web of certainties. This process, I take it, is non-deterministic.

In sum, the requirement of universality is not met in *OC*, as certainties can be culture-dependent. Specifiability also fails, as there is no theoretical way of delimiting the foundations from the superstructure. For independence, I propose a strong and a weak reading. Under the strong reading, basic beliefs may depend on nothing. This is surely not the case for *OC* as certainties are at least interdependent. A weak reading would require basic beliefs to only depend on other basic beliefs. That this view is found in *OC* is at least debatable as certainties can loose their certainty-status and in certain circumstances admit grounding by possibly less certain means. If we take seriously the time-dependence of the dependence relation, i.e. read 'x depends on y' as 'x depends on y *at time* t', we would again be faced with a strong (time dependent) relativism. Moreover, it is still questionable whether 'My name is M.B.' even if at times dependent on what my ID card says, is at those times less than certain, i.e. certainties may at times be dependent on non-certainties but may not thus be less than certain. This might also lead to a foundationalism in which knowledge itself is temporal, i.e. locutions such as 'I *knew* p at time t_1 but I do not know p at t_2' might be grammatical (with t_1 before t_2), as whatever 'knowing p' depended on at t_1 it may no longer depend on at t_2.[12] Thus, if a foundationalist at all, Wittgenstein is a fairly untraditional and at least moderately relativist foundationalist. He would also be an ingenious (relativistic) foundationalist in that he does not require that basic beliefs give rise to non-basic ones by means of logic, but rather constructs the certainties in a way that their justificatory force comes from within. We could, thus, *call* Wittgenstein a foundationalist. "But words matter," argues Williams [2007, pp. 49–50], and "we must recognize that foundationalism is more than the view that there are certainties of some kind or other, so that scepticism goes wrong *somehow*." In other words, strictly speaking Wittgenstein's anti-scepticism is not a form of foundationalism; loosely speaking it clearly is but that does not situate Wittgenstein within the foundationalist *tradition*.

[12] We have to be careful here since these constructions are quite grammatical with wh-complements: 'I knew *why* elephants have big ears a while ago but I don't know it now, I forgot.'

A last word is in order on the extent of the relativism found in *OC*. Williams [2007, p. 56] claims in *OC* Wittgenstein "repudiates relativism as strongly as he does scepticism." This is too strong, I think, as we have seen that Wittgenstein allows for at least some relativism. Discussing a tribe who believe they have been to the moon in their sleep (at a time where space travel was not as common place as today), he writes:

> What we believe depends on what we learn. We all believe that it isn't possible to get to the moon; but there might be people who believe that that is possible and that it sometimes happens. We say: these people do not know a lot that we know. And, let them be never so sure of their belief—they are wrong and we know it.
>
> If we compare our system of knowledge with theirs then theirs is evidently the poorer one by far. [OC, 286]

Here, we can argue, it becomes clear that there are sets of certainties that are superior to others. Not every view is as good as the next and, thus, the relativism found in *OC* is at least moderated.

1.2. THE TRANSCENDENTAL READING. Wittgenstein's certainties, being the "background against which [we] distinguish true and false," it seems make truth and falsity possible in the first place by being part (or the basis) of a language-game. This bridges over into the thought of Kant, whose "'transcendental logic' has to do with this 'possibility of truth and falsity"' [Brenner 2007, p. 122]. Kant, in his *Critique of Pure Reason* (henceforth *CPR*), is, among other things, concerned with conditions for thought; that is, in particular, his "[t]ranscendental logic deals with *a priori* and completely general principles which govern the construction of objects, and relate judgements to objects so that we may come to speak of *true* judgements" [Achourioti and van Lambalgen 2011, p. 9]:

> The part of transcendental logic that expounds the elements of the pure cognition of the understanding and the principles without which no object can be thought at all, is the transcendental analytic, and at the same time a logic of truth. For no cognition can contradict it without at the same time losing all content, i.e. all relation to any object, hence all truth. [CPR, A62-3/B87]

What Kant is looking for here are the *conditions* under which thought can relate to reality. That is to say, the aim is not necessarily to describe the *relation* between thought (or language) and reality itself. As Mounce [2007, p. 119] puts it:

> Classical metaphysics, let it be emphasized, is *not*, as is vulgarly supposed, an attempt to transcend the human condition by attaining an external standpoint. It is an attempt to illuminate that condition by making explicit what is only implicit in the condition itself.

Aiming for a Kantian interpretation of Wittgenstein's work, Brenner [2007, p. 125] further identifies Wittgenstein's *parts of speech* and *grammar*[13] in the *PI* with Kant's *transcendental forms of sensibility and understanding* and interprets Wittgenstein as attempting to solve Kant's *problem of philosophy*, namely "the nature of that correspondence between language, or thought, and reality that makes language *language* and thought *thought*" [Brenner 2007, p. 122].

The correspondence between language (or thought) and *reality* really falls in the realm of metaphysics rather than epistemology, but it is worthwhile looking at this briefly. For Wittgenstein, our chain of justification has to come to an end somewhere [cf. OC, 192]. This end, as we have seen, are the certainties. Now the question arises in how far *these* have to correspond to reality. We will not try to answer this question here.[14] The discussion of a Neo-Kantian interpretation of *OC* — and especially the point of divergence between Wittgenstein and Kant — will, however, cement the important dynamic dimension to certainties. Brenner [2007, p. 132] (citing Wittgenstein [OC, 395 and 397]) explicates the relationship between language and reality as follows:

> Our language, including our forms of reasoning, develops and expands through our interactions with the world. And conversely, the world-as-we-experience-it gets larger as we master more language. [...]
>
> It's a matter of the logic of language that chairs (for instance) are physical, moveable objects that can be caused to move by pushing them. It's a matter of fact that chairs exist and that people often move them around.

The distinction between the logic of language and the matters of fact is already present in Kant's oeuvre, there called *form* and *matter* (of appearances). Wittgenstein, so argues Brenner [2007, p. 132], adds to this the qualification "that the distinction [...] is not everywhere sharp." This becomes evident in the following, famous passage:

> [...] But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other. [OC, 97]

The river-bed, of course, is the end of what is justifiable — the certainties of our language-game —, the logic of our language; the waters are life (with

[13] To avoid confusion with typical linguistic uses of 'grammar' and 'parts of speech', 'grammar', as understood by Wittgenstein, is a much "wider — and more elusive — network of rules which determine what linguistic move is allowed as making sense, and what isn't" [Biletzki and Matar 2014, Sec. 3.6]. The *part of speech* a word is assigned to "determine[s] the *kind of use* of the word" [PG, 25] and signals what the grammar of a word is. For example, color words, as parts of speech, have a certain applicability (or grammar) rendering such constructions as 'the color of the number 3' nonsensical.

[14] I refer the reader interested in the realism/anti-realism debate with respect to *OC* to Mounce [2007], arguing for the (rather controversial) position that this late Wittgenstein is a realist, and Brenner [2007], arguing the contrary.

language), the empirical statements about the world. Even more is suggested by this metaphor and its continuation:

> And the bank of that river consists partly of hard rock, subject to no alteration or only to an imperceptible one, partly of sand, which now in one place now in another gets washed away, or deposited. [OC, 99]

Not only is there no sharp distinction between what stands fast and what flows freely; what flows freely can become fast-standing and what stood fast once may begin to flow. This is certainly a departure from Kant: "With the introduction of these *relatively* stable 'forms of life', there is clearly no room left for anything like Kant's *absolutely* stable 'transcendental forms' " [Brenner 2007, p. 125, my emphasis].

The metaphor of the riverbed in flux poses the question whether *logic* itself is unchangeable [cf. *ibid.*, p. 133]. If *everything* may be subject to change, "what becomes of logic now? [...] doesn't logic altogether disappear?" [PI, 108]. In *OC* itself there seems to be nothing suggesting a negative answer to this question. Indeed, if we take Wittgenstein's reflection on the rigour of mathematics in *OC* as any indication, then it rather seems as though logic is not in any way less susceptible to change than any other deeply entrenched certainties. While Wittgenstein [OC, 657] suggests that "propositions of mathematics might be said to be fossilized" he does admit in OC, 563 that there is mathematical knowledge, gained by proof but where there is knowledge there can be doubt (cf. e.g. section 1.3). We will take this issue back up in section 6. It is the wide-spread dynamic dimension of certainties, I argue, that is an important facet to carry over from this metaphysical discussion to our epistemological approach.

1.3. THE THERAPEUTIC READING. As has been done by the New Wittgensteinians[15] with all of Wittgenstein's work, also *OC* has a reading under which it is merely therapeutic in the sense that it is to treat our linguistic confusions.

> For the Therapeutic reader, Wittgensteins talk of conditions, foundations and scaffoldings is, like talk of ladders, something which we should regard as either emetic or homeopathic — either something we swallow in order to ruminate and reject it; or a sufficient dose of the poison to cure us of the poison. [Moyal-Sharrock and Brenner 2007, p. 4]

In sum, this reading takes *OC* to be aimed at dissolving the linguistic confusions the sceptical philosopher and Moore have brought to the table. In gearing up for a therapeutic reading of *OC*, Minar [2007, p. 261] explains:

[15]The *New Wittgensteinians* "have in common an understanding of Wittgenstein as aspiring, not to advance metaphysical theories, but rather help us work ourselves out of confusions we become entangled with when philosophizing" [Crary and Read 2002, p. 1], that is, it is the view that "Wittgenstein's primary aim in philosophy is [...] a *therapeutic* one" [*ibid.*].

> Wittgenstein's aim in *On Certainty* is *not* to demonstrate that Moore's common-sense realism and the sceptic's challenge to it are meaningless. He labours to bring to the fore each party's underlying commitments and to raise questions about the reasons we might have for accepting or rejecting them. [...] The picture of scepticism to which our reading *On Certainty* leads us may seem incomplete, inaccurate or unfair, but [...] Wittgenstein's procedures anticipate this concern: He is striving to create a context in which we will be drawn to give voice to the sceptical moments in our own thinking; and he then wants to force us to account for our sceptical impulses, thus placing a burden on *us* to explain how any sceptical tendencies we may still harbour have been misrepresented.

'I know', for Wittgenstein, is appropriate only in certain circumstances — where there is a "possibility of satisfying oneself" [OC, 3] and where it can be "*shewn* that no mistake was possible" [OC, 15] — and "must admit of being established objectively" [OC, 16]. Usually, in these very same circumstances 'I doubt' is in order as well.

Moore, however, only lists (or enumerates) common-sense utterances that he claims to know. He fails to show that he *does know* these claims and, in fact, fails to "[determine] what showing that he does know would be" [Minar 2007, p. 262]. All Moore can go by is his own, first-person authority which, as Wittgenstein points out, is overlooking the form 'I thought I knew' [cf. OC, 12]. Does this mean Moore's utterances are nonsense? Consider the following passage:

> [...] 'I know that there's a sick man lying here', used in an *unsuitable* situation, seems not to be nonsense but rather seems matter-of-course, only because one can fairly easily imagine a situation to fit it, and one thinks that the words 'I know that...' are always in place where there is no doubt, and hence even where the expression of doubt would be unintelligible. [OC, 10]

Minar, citing OC, 348, interprets Wittgenstein as implicitly conveying here that Moore's utterance 'I know here is a hand' is not nonsensical because it is an inappropriate use of 'I know' but rather because 'here is a hand' is not (not yet, at least) an object that meaningfully connects with 'I know', *in the sense* relevant for Moore's argument to go through:

> The nonsensicality of Moore's utterances is not a function of a misfit between pre-existing propositions and their present contexts. Rather, their 'meaning is not *determined* by the situation, yet stands in need of such determination' [OC, 348]. We should not rush to assign a particular epistemic status to the sentences Moore recites, then, because there are as yet no items to which such a status can meaningfully be assigned. [Minar 2007, p. 262]

On the one hand, Moore cannot use 'I know' as he does not have the means to objectively show that he *does* know, i.e. rule out any possible mistake. On the other hand, as Wittgenstein [OC, 17] points out, we have no clear idea of what he could possibly be mistaken about: "What would a mistake here be like? And have I any *clear* idea of it?"

Turning now to the sceptic, we can see that the burden is on him as much as it is on Moore to explain what "having proper grounds amounts to" in this case [Minar 2007, p. 262].

This observation leads us to realize some of the things the sceptic is committed to. "'Doubting the existence of the external world' does not mean for example doubting the existence of a planet, which later observations proved to exist" [OC, 20]. Rather, the doubt the sceptic raises is not a *practical* doubt — he is not doubting the existence of a particular object — but a further, non-practical doubt. What he seems to target is our framework of dealing with claims about all physical objects. "That is to say, he questions whether the language-game within which the notion of 'proper ground' operates is adequate. But adequate to, or for, what?" Minar [Cf. 2007, pp. 262–3]. The therapy lies in asking the sceptic what he is doing. What is he getting at here? What is this non-practical doubt?

> It may be, first, that for all we know it is never possible for the trustworthy man to check *thoroughly* enough about the existence of a particular physical object. Or it may be, second, that the sheer possibility that there are no physical objects is being pondered [...]. [*ibid.*, p. 263]

On the first reading, the sceptic's non-practical doubt is concerned with the end of our procedure of justification; i.e. when is enough *really enough*? This, so Wittgenstein, according to Minar [*ibid.*], is rather unclear:

> The idealist's [or sceptic's, cf. OC, 37] question would be something like: "What right have I not to doubt the existence of my hands?" (And to that the answer can't be: I *know* that they exist.) But someone who asks such a question is overlooking the fact that a doubt about existence only works in a language-game. Hence, that we should first have to ask: what would such a doubt be like?, and don't understand this straight off. [OC, 24]

That a non-practical doubt, be it of the first or second kind, needs to be expressed within a language-game is hardly a satisfying response. "At most," argues Minar [2007, pp. 263-4], "it shows that the first alternative above, [...] calls for caution, not sceptical suspension of belief." That is to say, a further doubt of the first kind does not necessarily drive us into scepticism proper, but rather focuses our attention on the role the very notion of *making sure* plays in our language-game. The second type of further doubt — where the question "has become the issue of the probity of *any* of my ways of inquiring for telling me *anything* about the existence of a mind-independent external world" [*ibid.*, p. 264] —, however, is not addressed at all. In either case, Wittgenstein does not seem satisfied with the clarity of the question (or doubt) raised by the sceptic. Clearly, the sceptic cannot doubt every single judgement, for that would leave him outside

of the realm of judging altogether; "If you tried to doubt everything you would not get as far as doubting anything" [OC, 115] for then even the initial doubt itself would be doubted. In certain circumstance, certain things have to be beyond doubt, mistakes have to be excluded.

The therapy continues, as Wittgenstein [OC, 26] supplies the sceptic with a follow-up question: "But can it be seen from a *rule* what circumstances logically exclude a mistake in the employment of rules of calculation?" As we have seen in subsection 1.1, such a rule cannot be formulated.

> What is 'learning a rule'? *This*.
>
> What is 'making a mistake in applying it'? *This*. And what is pointed to here is something indeterminate.

Minar [2007, p. 265] argues that Wittgenstein not only underlines in this passage that particular judgements (or mistakes), as it were, draw out the game to show us how to partake in it and that detaching ourselves from these judgements will leave it unclear whether we are still playing the game. Moreover, Minar [*ibid.*] interprets Wittgenstein, "[w]hen he says that something 'indeterminate' lies at the heart of the practice of following a rule, he means that at some point there will be nothing deeper to appeal to" thus distancing himself from the sceptic. The sceptic asks too much. He questions the underlying structure of our practices of knowledge-claims, questions how knowledge-claims are possible without a clear picture of the substratum. Under this reading, Wittgenstein denies that we *can* have a clear picture of these practices and, thus, "no global question of adequacy of these proceedings has been given a foothold" [*ibid.*].

> [. . .] But the most important thing is: The rule is not needed. Nothing is lacking. We do calculate according to a rule, and that is enough. [OC, 46]
>
> [. . .] Forget this transcendent certainty, which is connected with your concept of spirit. [OC, 47]

Both Moore and the sceptic are guilty of assuming there is a transcendental structure according to which our judgements are assigned epistemic statuses [cf. Minar 2007, p. 266].

> Clearly, [Wittgenstein] wants to say that both Moorean realist and sceptic purport to hold 'there are physical objects' up for something like empirical support, while at the same time ruling out the *possibility*, the very coherence of the idea, of such support prior to the raising of any actual doubts. [. . .] it will be unclear from what point of view the ostensibly (but dubiously) empirical proposition 'there are external objects' can be challenged *or* supported. It has not been determined what the *it* that would be challenged or supported would *be*. [*ibid.*, p. 268]

This reading, to be sure, poses a problem for our project at hand. How do we formalize epistemic notions in a setting where they are denied a determinate underlying structure? Wittgenstein, according to Minar's reading, is not constructing a theoretical framework but a therapeutic context. A context in which we clarify the sceptic's underlying commitments and at

the same time one in which it becomes less clear and easy to even understand him.

But Wittgenstein, in *OC*, does not (or not just) lay out a therapeutic context. Some remarks make it clear he is getting at a connection between our ways of life, our games and the world. "This game proves its worth" [OC, 474] and when comparing the systems of certainties guiding our games, we can say that one "is evidently the poorer one by far" [OC, 286].[16] Some systems of certainties, some language-games are better than others because they better *fit*[17] the way the world actually is; this fit proves their worth and "may be the cause of its being played, but it is not the ground" [OC, 474]. By admitting such a connection, however lose or vague it may be, Wittgenstein is, after all, advancing a positive metaphysical theory instead of a mere therapy of our confusions.

1.4. THE EPISTEMIC READING. The epistemic reading of *OC* is primarily concerned with the question whether certainties are *epistemic* in nature at all. This is surely left as an open question by the therapeutic reading, according to which Wittgenstein offers insights into the underlying commitments of our knowledge-claims and doubts. Morawetz [2007, p. 165], however, warns us about a fallacious inference form the fact that knowledge-claims can shed light on the concept of knowledge. The fallacy would be to infer from this that there is a one-to-one correspondence between knowing and being able to express this.

This limit of expressible knowledge, so Morawetz, is evident in the following passage:

> I am sitting with a philosopher in the garden; he says again and again "I know that that's a tree", pointing to a tree that is near us. Someone else arrives and hears this, and I tell him: "This fellow isn't insane. We are only doing philosophy." [OC, 467]

The point is that knowledge-claims, when uttered in the wrong context, seem *queer* [cf. OC, 553] — that is not to say they are unintelligible (as far as we can imagine a fitting context), but merely odd. For example, if we are having a casual conversation about, say, music and amidst this conversation I suddenly utter 'I know my name is M.B.', you would be frazzled. There is, in the context of our conversation, no place for such a claim. It is inappropriate to utter in most normal circumstances, even though it is undoubtedly correct that that is my name and that, in a fitting context, I can give grounds for the claim and, thus, in those circumstances be said to know it. A context in which such a claim may make sense, then, has to again involve "a way of making sure" [cf. Morawetz 2007, pp. 165-7].[18]

Now that we have established the context-dependence of knowledge-*claims*, the question, as Morawetz [2007, p. 167] points out, 'When can I be

[16] Cf. also the end of section 1.1, where we discuss the extent of the relativism in *OC*.

[17] I purposefully avoid using the term 'corresponds' here.

[18] This observation — that knowledge-claims only make sense in a given context — leaves the Moorean realist as well as the sceptic in a situation similar to the one the therapeutic reading has left them in:

said to know p?' becomes three-fold. It could be the question (i) When is it *true* that I know p? (ii) When is it *appropriate of me* to claim that I know p? and (iii) When is it *appropriate of others* to attribute to me that I know p?

Let us begin by addressing (i):

> An inner experience cannot shew me that I *know* something. [OC, 569]
>
> The utterance "I know..." can only have its meaning in connection with the other evidence of my 'knowing'. [OC, 432]
>
> But don't I use the words "I know that..." to say that I am in a certain state, whereas the mere assertion "that is a..." does not say this? And yet one often does reply to such an assertion by asking "how do you know?" [...] [OC, 588]

Knowledge, thus, is a state in which I bear a certain relation to a fact. The activity of judging whether I know only comes into play when I am asked to give grounds for my knowledge — to justify my claim — and is connected to my state of knowledge logically rather than causally. That is to say, my being in a state of knowledge does not *cause* me to give grounds, but reveals itself in my giving of grounds. "My knowing p is not an inner state of being that causes me to act in certain ways, for example to give grounds, but rather it is manifested when I act in such ways" [Morawetz 2007, pp. 168-9]. If, however, someone were to doubt that I know p, they may doubt p itself or the fact that I am among those who know p. The difference between uttering 'p' and 'I know p' vanishes in presence of the speaker:

> [...] 'I know' has a meaning only when it is uttered by a person. But, given that, it is a matter of indifference whether what is uttered is 'I know...' or 'That is...'. [OC, 588]

Turning now to questions (ii) and (iii), we address the distinction between first- and third-person knowledge claims.

> If someone says, "I know that that's a tree" I may answer: "Yes, that is a sentence. An English sentence. And what is it supposed to be doing?" Suppose he replies: "I just wanted to remind myself that I *know* things like that"'?— [OC, 352]
>
> But suppose he said "I want to make a logical observation"? — If a forester goes into a wood with his men and says "*This* tree has got to be cut down, and *this* one and *this* one"— what if he then observes "I *know* that that's a tree"?—But might not *I* say of the forester "He *knows* that

---

'I know there are physical objects' might be an utterance of a peculiar kind. Can we ever make sense of it providing a context? Under virtually all circumstances, the claim would raise eyebrows and prompt puzzlement. [Morawetz 2007, p. 167]

> that's a tree—he doesn't examine it, or order his men to examine it"? [OC, 353]

For example, when Ada says 'I know p', she signals that she can give evidence for p. If true, it implies that p is true and that Ada bears a certain relation to p (that of knowing p). When Bob says 'She [Ada] knows p', he signals no such commitment (giving evidence for p). If true, the claim implies that p but not that Bob bears the knowledge relation to p. Rather, Bob must give reasons for his thinking *Ada* knows p, i.e. he signals he has reason to believe *she* can give reasons for p, bears the relation of knowledge to p. Bob's claim is then dependent on Ada's performance. Both of these claims are subject to fail when the proper reasons cannot be provided by *Ada* [cf. Morawetz 2007, pp. 169–70], but the third person claim can also fail if Bob cannot provide reasons for thinking Ada can provide reasons.

> Whether I *know* something depends on whether the evidence backs me up or contradicts me. For to say one knows one has a pain means nothing. [OC, 504]

The last sentence in the above quote circles us back to the appropriateness of knowledge-claims — knowledge-claims presuppose the possibility, at least in principle, to be backed by evidence. Then, in principle, in all these cases of appropriate uses of 'I know' there is a possibility of being wrong, a possibility of *doubt*.[19] If I say 'I know I had a pain in my shoulder earlier' I have inappropriately used 'I know' since I cannot possibly doubt it. Moreover, it seems that, as 'I know. . .' describes an internal state just as 'I had a. . .' does, the use of 'I know' here is simply superfluous [cf. Morawetz 2007, p. 170].

> The wrong use made by Moore of the proposition "I know . . ." lies in his regarding it as an utterance as little subject to doubt as "I am in pain". And since from "I know it is so" there follows "It is so", then the latter can't be doubted either. [OC, 178]

Now since *my* utterance 'p' would commit me to giving grounds for p, what does prefixing 'I know' add to the conversation? That 'I know p' and 'p' are not, in fact, equivalent comes out when considering their negations:

> Even when one can replace "I know" by "It is. . ." still one cannot replace the negation of the one by the negation of the other. With "I don't know. . ." a new element enters our language-games. [OC, 593]

As Morawetz [2007, p. 171] correctly points out, even though 'I know p' and 'p' are interchangeable in many contexts, "only 'I know p' explicitly states both the truth of p and the fact that I bear a particular relationship to p. 'I don't know p' denies that that relationship exists." Wittgenstein compares

[19] Wittgenstein, in [PI, 244–271], has previously discussed such matters in the so-called *private language argument*. There [PI, 243], Wittgenstein asks us to imagine a language in which "individual words [. . .] are to refer to what can only be known to the person speaking; to his immediate private sensations" How to interpret the argument is contentious [cf. Candlish and Wrisley 2014], but it is clear that such *knowledge* (if this term even applies) cannot be expressed in public.

this particular relationship to that evoked by 'I see': "'I know' is supposed to express a relation, not between me and the sense of a proposition (like 'I believe') but between me and a fact" [OC, 90].

The objectivity conveyed by 'I know' separates it from such locutions as 'I am certain', as the latter express a subjective attitude towards a fact [cf. Morawetz 2007, p. 173]. Now knowledge-claims can be withdrawn — "I though I knew" but never *did know* — in light of new evidence. But they cannot *all* fail at once [cf. OC, 506, 614–617]. The point is "that claims to know p may assure others of the truth of p but never *insure* truth" as they can be deceitful or mistaken [cf. Morawetz 2007, p. 177].

This fallibility of knowledge-claims may prompt people to express themselves more carefully: "Suppose it were forbidden to say "I know" and only allowed to say "I believe I know"?" [OC, 366]. Wittgenstein [OC, pp. 483–6] makes it clear that 'I know' and 'I believe' play a very different role in our language-game — the latter not committing the speaker to be able to give adequate or convincing grounds. The former, however, signalling that, if the speaker were wrong, uttering it would have been misconduct [cf. OC, 367]:

> Isn't it the purpose of construing a word like "know" analogously to "believe" that then opprobrium attaches to the statement "I know" if the person who makes it is wrong?
>
> As a result a mistake becomes something forbidden. [OC, 367]

Knowledge-claims, while fallible, constitute a wrong usage of 'I know' if they turn out to be false. To use 'I know' correctly is to make sure that no mistake was possible in my grounding my claim. Under this assumption, knowledge has to be taken to be (positively) introspective, i.e. 'I know' implies 'I know I know':

> "If I know something, then I also know that I know it, etc." amounts to: "I know that" means "I am incapable of being wrong about that". But whether I am so must admit of being established objectively. [OC, 16]

Some propositions "of the form of empirical propositions" [OC, 96], however, may not be taken as objects of knowledge claims; they only *look* like empirical propositions but are "not testable in any context" [Morawetz 2007, p. 180]; we may call them *methodological* propositions [cf. OC, 318]. That is not to say that methodological propositions and propositions that are suitable to be complement clauses of knowledge-claims exhaust the types of propositions alluded to in *OC*. Morawetz [2007, pp. 180–2][20] identifies at least four types of propositions: (i) methodological propositions that can *never* be part of knowledge claims (e.g. 'physical objects exist', which cannot be empirically verified); (ii) propositions that can *rarely* be part of knowledge claims (e.g. 'my name is M.B.', which is checkable but rarely checked); (iii) propositions that are easily revisable but only in highly unusual circumstances that would possibly result in us questioning our own sanity (e.g. 'in continental Europe we drive on the right side of the road',

[20] It is important to note that Wittgenstein himself "does not discriminate between the first and second category" [*ibid.*, p. 180].

which is checkable and *may* turn out to be false, but only if 300 million people turn out to be delusional); (iv) propositions which are open for discussion and the main class of complement-clauses of appropriate knowledge-claims (e.g. 'cows have a stomach with four compartments', as uttered by a biologist).

Under this reading, methodological propositions of type (i) clearly fall under the class of certainties. Their unverifiability makes them unrevisable by rational means and, thus, part of the hard rock bottom of the riverbed alluded to in section 1.3. In a sense, they are to be taken like propositions of logic, they belong to the logic of our language-game. That is not to say they cannot change, but "such a 'revision' would amount to annihilation of all yardsticks" [OC, 492], i.e. a change of my *method of inquiry*. Again, nothing *forces* one to give up such propositions and for all intents an purposes they stand fast, but there are *objectively* better sets of such propositions (cf. section 1.1) and I may change them after all on grounds beyond my method.

On top of this layer, the propositions of type (iii) make up more of the riverbed; this part is subject to change within the language-game though rarely does change. Type (iv) propositions make up the waters of the river itself. Where exactly the type (ii) propositions fit into this picture is not quite as clear; perhaps between the type (i) and type (iii) propositions or, perhaps more likely, at the bottom of the riverbed along with the type (i) propositions [cf. footnote 20]. To be clear, propositions of types (i)–(iii) are all certainly reflected in our (inter)actions — they are "*in deed* not doubted" [OC, 342].

As we have seen in sections 1.1 and 1.3, the categorization of these four types of objects as being *propositions* is not uncontentious. For example, we could read Wittgenstein as stressing the meaninglessness of Moore's utterance 'here is a hand'. While it is certainly true, it is not appropriate to express this and, under this reading, down-right meaningless. But propositions, as commonly understood,[21] are precisely the *meanings* of sentences. Thus, if such utterances as 'here is a hand' are meaningless, there is no proposition attached to it, and, hence, hinges or certainties are not propositions [cf. Pritchard 2007, p. 199]. Pritchard [*ibid.*] argues that, if we grant that they are, in fact, propositions, "it ceases to be obvious that *On Certainty* is offering an *epistemological* response to the sceptic." The response, as it were, could be seen as being merely semantic — what is being expressed *simply cannot be said*, it goes against the logic of our language.

2. Aspects

In this section, I want to collect some of the remarks we have already encountered in a more perspicuous way. We will look at certain aspects of *OC* that deserve further investigation and a closer reading of the original remarks.

2.1. HINGES. Pritchard [2007, pp. 194–200] offers what he calls "the minimal reading" of hinges, which we will take as a basis for this section. There,

[21] A discussion of what exactly a *proposition* is — in general or for Wittgenstein — is far beyond this paper. Cf., however, McGrath 2014.

Pritchard argues that the fact that hinges are not "legitimately prone to coherent doubt" is not merely due to the fact that they are "*in deed* not doubted" [cf. OC, 342], but rather that they have a normative dimension to them; "we *ought* not to doubt them" [Pritchard 2007, p. 195]. Doubt, like knowledge, presupposes a ground that is more certain than the proposition doubted (or which is claimed to be known). Nothing, however, is more certain than a hinge and "thus [hinges] are exempt from epistemic evaluation of both a positive *and* a negative sort" [*ibid.*, p. 197]; thus, "an assertion of a first-person knowledge-claim [or doubt] regarding a hinge proposition is always *conversationally* inappropriate" [*ibid.*, p. 198, my emphasis]. Hinges, *by their nature* [cf. *ibid.*, p. 198], are not justifiable — and hence not justified or evidentially supported — by anything but themselves.

This much about hinges is quite uncontentious [cf. *ibid.*]. The mere impropriety of claiming to know (or asserting a doubt about) a hinge, however, is not the strongest reading *OC* admits. As we have seen, such claims may be viewed as incoherent, nonsensical or downright meaningless. This returns us to our discussion towards the end of section 1.4: Are hinges propositions? Pritchard [2007, p. 199] argues that in light of the minimal reading "hinge 'propositions' are, it seems, *beliefs* rather than propositions" since the hinge-status of these objects can change with the context, i.e. one and the same object can at one time be a hinge and at another an ordinary empirical proposition. 'Beliefs' here, will have to be taken as an attitude (of confidence) one has towards a proposition, as 'belief', *per se* comes with degrees of *assurance* [cf. OC, 66] and thus admits giving evidence, doubt, etc. *Propositions*, understood as meanings of sentences, are abstract entities and not subject to change — although our attitudes towards them may change frequently.

We will, in the following, take certainty to be a form (the strongest form) of a belief-like (subjective) attitude we can have towards a proposition. We will, at the same time, however, take the hinge propositions to be *propositions* and thereby let them be the sort of object that can be taken as the complement-clause of a belief- or certainty-assertion.[22] When we talk of hinges think of *sets* of propositions. Propositions can at one time be part of the set, at another time they may not be; their status as certainties is identified with membership in the set.

Taken as a sort of belief, certainty is an utterly *subjective* notion.

> It would be correct to say: "I believe…" has subjective truth; but "I know…" not. [OC, 179]
>
> Or again "I believe…" is an 'expression', but not "I know…". [OC, 180]

These expressions of belief come in *degrees*, the maximum value of which is certainty.

> Anyone who says, with Moore, that he knows that so and so…—gives the degree of certainty that something has for

[22] The *conversational* inappropriateness of knowledge-claims about such propositions can then be handled by a normative or pragmatic principle such as, say, "It is inappropriate for me to utter 'I know p' if I am certain of p."

> him. And it is important that this degree has a maximum value. [OC, 386]

But not just beliefs come in degrees, certainties do as well. We recall that Morawetz identified four types of propositions that can be found in *OC*, that stand-fast in different degrees. Under this reading, then, we have at least two levels of certainties: at level 1 we find the methodological propositions of type (i), which have the form of empirical propositions but do not admit any testing. At level 2 we find propositions of type (ii), which are in principle verifiable but rarely stand in need of verification, as well as propositions of type (iii), which are entrenched in our actions and take highly unusual circumstances to be subject of doubt. Perhaps we need to allow for subdivisions of these levels but the main difference is the fact that level 1 certainties cannot be (rationally) doubted in *any* circumstances, while level 2 propositions admit doubt but only in rare and unusual circumstances in which almost everything is up for grabs. At any level of certainties, we can go from subjective certainty to *objective* certainty.

> [...] But when is something objectively certain? When a mistake is not possible. But what kind of possibility is that? Mustn't mistake be *logically* excluded? [OC, 194]
>
> "I have compelling grounds for my certitude." These grounds make the certitude objective. [OC, 270]
>
> What is a telling ground for something is not anything *I* decide. [OC, 271]

Here, the rationale between the divisions of the levels comes to light. The first quote above explains how level 1 certainties can be objectively certain — type (i) propositions are methodological and they are part of our *logic of inquiry*, our logic of language and, thus, *logically* excluded from doubt. Level 2 certainties do admit grounding (albeit very seldom appropriate), so for them to be established as objectively certain one has to give *objective* grounds. In these cases then, knowledge-claims of these certainties are in order.[23]

On top of these levels of certainties we find, to varying, less-than-certain degrees of belief, propositions of type (iv), which are common, everyday empirical propositions. These degrees of belief may be viewed as *degrees of assurance* or *conviction* [cf. OC, 66]. Being the waters of the river, to stick with Wittgenstein's metaphor, these propositions are prone to change their degree of assurance based on rational argument and may, in certain cases, achieve certainty-status. Likewise, level 2 certainties may, in certain circumstances, become unhinged and lose their certainty-status. Level 1 certainties cannot change so easily though they are, as we have seen in subsection 1.1, not in general *universal* (we will go into this deeper in subsection 2.2).

All the talk of different levels should not, however, let the picture of the inverted pyramid of the foundationalist resurface. The membership of a given proposition to one level (or type) or another is not, in general, decidable. While we can give examples, as we did above, and point to certain

[23] At level 2, "I know = I am familiar with it as a certainty" [OC, 273].

characteristics (as we also did above) we cannot, without reference to normal circumstances, decide what level a given proposition belongs to [cf. OC, 27]. Most likely, what counts as a 'normal circumstance' is subject to change as much as anything and what I here take to be normal circumstances might not be normal circumstances to everyone (e.g. if I were an Alzheimer's patient, 'my name is M.B.' would be uncertain and in need of verification in most circumstances normal (?) *for me*).

Let us begin at the bottom. To decide for any proposition, whether it is of type (i), (ii) or (iii) is not possible as the inability of describing *ordinary* circumstances under which, e.g., propositions of type (ii) or (iii) are not subject to doubt, follows from the our inability to express a rule for this procedure.[24] Hence, it is undecidable whether a given proposition is of type (i), (ii) or (iii). The capability of the propositions of type (iv) to enter level 2 and level 2 proposition's capability to be, at times, of type (iv), finally, blurs the line between certainties and degrees of assurance.

The point of splitting beliefs and certainties into different levels is to show that there are different levels of entrenchment to the objects of our beliefs (and, hence, to the things that stand fast for us). We should, furthermore, not forget that these levels, as non-delimited as they are, form a single, indivisible structure of convictions. (Of course, the structure may change over time.) "Not that I could describe the system of these convictions. Yet my convictions do form a system, a structure" [OC, 102].

> All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments: no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life. [OC, 105]
>
> It is quite sure that motor cars don't grow out of the earth. We feel that if someone could believe the contrary he could believe *everything* that we say is untrue, and could question everything that we hold to be sure.
>
> But how does this *one* belief hang together with all the rest? We should like to say that someone who could believe that does not accept our whole system of verification.
>
> This system is something that a human being acquires by means of observation and instruction. I intentionally do not say "learns". [OC, 279]

2.2. CONFLICTING FRAMEWORKS. We now briefly turn to the universality of hinges. As the quote from OC, 279 illuminates, the systems mentioned in subsection 2.1 are "systems of verification", i.e. "the inherited background against which I distinguish between true and false" [OC, 94]. However, these systems are not universal and more tightly knit than the exposition of

[24] Cf. also subsection 1.1 and the discussion on specifiability.

subsection 2.1 might suggest. Recall the story about Moore's disagreement with a king who believes the world to have existed only since his birth.[25]

From the example of the king it becomes clear that it is easily imaginable that someone might disagree with us at the level of certainties. There is more to it than that. Clearly, such a statement as 'motor cars don't grow out of the earth' and arguably 'the earth has existed before my birth' are not level 1 certainties. Rather, they should belong to level 2 as there are imaginable circumstances in which I would have reason to doubt them. (Say, if I saw a car grow on a tree.) If anything, such a proposition is of type (iii), as it would make me question my sanity if I witnessed an event revising my belief in it.

What this shows, then, is that, at any level of entrenchment, certainties are of importance to our system of verification. The likelihood of them behaving dynamically may still depend on how deeply entrenched they are. The type (iii) proposition 'motor cars don't grow out of the earth' has an effect on the whole system. Everything in my system can be deemed false by someone who allows for doubt in this particular instance [cf. OC, 279]. Disagreement in but one instance can mean disagreement over the whole system. Yet still, it may be easier for me to give up *that* belief than, say, my belief in physical objects. What this further shows is the intrinsic connection between certainty and knowledge, as the notion of *proper grounds* that gives meaning to the notion of knowledge rests on the propositions that are certain, depends on our *method*.

There might be no way to *rationally* convince someone who doubted that cars don't grow on trees of the contrary, as he might not accept any of the grounds we can give to convince him otherwise — he would not accept our form of reasoning, in general. (Indeed, by my own admission, in any circumstance in which the other's doubt is in order, I would likely be insane.) Only *persuasion* can make the other realize the ungroundedness of his doubt, and this persuasion must be reached by extra-logical means (read: not by the logic of the language of one of the participants). Wittgenstein even mentions such criteria:

> [...] Remember that one is sometimes convinced of the *correctness* of a view by its *simplicity* or *symmetry*, i.e., these are what induce one to go over to this point of view. One then simply says something like: "*That's* how it must be." [OC, 92]

As mentioned in subsection 1.1, say I let myself be persuaded that, in fact, cars *can* grow on trees. Perhaps because it is a more simple explanation for the existence of cars than my belief in the complicated ongoings in car factories. My whole system would break down. I would now accept that cars grow on trees but what constitutes the web of certainties surrounding this conviction would still be indeterminate.

Rational communication (or interaction) with others thus relies on some intersection of the respective sets of subjective certainties. Communication between parties with disjoint or sufficiently disagreeing sets of certainties

[25] Cf. also subsection 1.1.

is not guided by logic. These disagreements, I would argue, are to be found mostly among the propositions of types (ii) and (iii), as type (iv) propositions are up for grabs in most situations and type (i) propositions can in no circumstances be doubted. If the possibility of disagreement about type (i) propositions is allowed for, I believe it would simply mean the breakdown of *communication* altogether. It would mean the logics of the respective disagreeing parties contradict each other. We should stress that we should distance ourselves from the idea that somehow type (iii) propositions are thus *less important* than, say, type (i) propositions as doubting one of them, too, can have far-reaching ramifications and lead to a breakdown in rational argument or collapse my whole system.

2.3. FIRST- AND THIRD-PERSON KNOWLEDGE CLAIMS. All these subjective certainties form "the inherited background against which *I* distinguish between true and false" [OC, 94, my emphasis]. They make up the system in which I define what counts as a ground and what counts as having made a mistake. It is thus, as mentioned, tightly connected to the notion of knowledge. But knowledge, for Wittgenstein, is not subjective.

> [...] Giving the assurance "I know" doesn't suffice. For it is after all only an assurance that I can't be making a mistake, and it needs to be *objectively* established that I am not making a mistake about *that*. [OC, 15]

Thus knowledge does not rest solely on *my* system of certainties, but also on an overarching game, i.e. an interaction with others where I give grounds. Indeed, knowledge depends on facts, "on whether the evidence backs me up or contradicts me" [OC, 504], but also on what constitutes a proper ground. We have to strongly distinguish between my subjective perspective comprised of my certainties and the objective perspective comprised of facts. Facts, however, are not solely what make up the objective certainties — communities can establish objective certainties as well. *My* certainties, however, cannot be wholly subjective; there has to be some objective dimension to them as Wittgenstein denies the possibility of a *private language* (cf. footnote 19). We have to especially distinguish subjective and objective perspective when *claims* of knowledge are involved.

> For it is not as though the proposition "It is so" could be inferred from someone else's utterance: "I know it is so". Nor from the utterance together with its not being a lie. — But can't I infer "It is so" from my own utterance "I know etc."? Yes; and also "There is a hand there" follows from the proposition "He knows that there's a hand there". But from his utterance "I know..." it does not follow that he does know it.

I cannot infer anything from utterances alone, only from utterances *taken as true*. If I have established the truth of my utterance 'I know p', I can infer p; likewise, if I take as true the utterance 'a knows p', I can infer p. Knowledge-claims themselves are fallible, be they first- or third-person claims. From my own claim to know p, however, I can infer p (as I would be

sure I have grounds for p); from someone else's claim to know p I cannot, as the other person has not yet given grounds.

Both kinds of claims can be viewed from an internal and an external point of view. The internal view can show us when someone is (sincerely) inclined to utter that one knows or someone else knows something. On the external view we can judge whether these claims are, in fact, true and thus constitute genuine knowledge. There is more. The notion of 'proper ground' may, indeed, be culture-specific. Recall, there are two elements to verifying knowledge-claims: first, 'I know' signals that I can give proper grounds and, second, these proper grounds need to be backed up. The community in which I utter a knowledge-claim has a hand in assessing whether it constitutes a legitimate case of knowledge. The collective deems my ground a *proper* ground (or not). This ground, then, has to live up to the facts (be they in accord with reality or, again, culture).

2.4. CERTAINTY, DOUBT, KNOWLEDGE, AND BELIEF. We want to now take a closer look at the most important epistemic notions — certainty, doubt, belief, and knowledge — and how they interact. As mentioned, we take certainties to be belief-like, i.e. propositional attitudes. We have also said that all instances of knowledge are instances of belief. Wittgenstein [OC, 308], however, stresses that "'[k]nowledge' and 'certainty' belong to different *categories*." It seems we have made a mistake in supposing that knowledge and certainty fall into the same *epistemic* category, viz. in the category of belief. However, as the surrounding passages [OC, 306–8] suggest, Wittgenstein is concerned here with *grammatical* categories, i.e. the function of 'knowledge' and 'certainty' in our language-game is categorically different. 'I know', as mentioned, expresses that I am in a certain relation to a fact; 'I believe' expresses a relation between me and the sense of a proposition [cf. OC, 90]. That is to say, 'I know dogs bark' establishes that dogs bark as a *fact* and my being in a relation to that fact, while 'I believe dogs bark' expresses that I have a certain *attitude* towards the *proposition* 'dogs bark'. 'I am certain' expresses my subjective conviction towards a proposition. It should be noted that, for all intents and purposes, the "difference between the concept of 'knowing' and the concept of 'being certain' isn't of any great importance at all, except where "I know" is meant to mean: I *can't* be wrong" [OC, 8]. In practice, they are used synonymously, but they do, in the setting of *OC*, perform different tasks.

For one, knowledge-claims about certainties are inappropriate. The targets of 'I know' utterances are empirical propositions that are subject to doubt; the targets of expressions of certainty are only seemingly empirical propositions but there is no room for doubt. Varying degrees of conviction, however, range over both target groups and members of one group can become members of the other and vice versa. Thus, it seems quite acceptable to group them together under the concept of 'belief-like', or 'propositional attitude'.

The strongest form of conviction towards a proposition is certainty. Things I can be said to know, thus, have to come in different strengths of conviction and things that I can appropriately claim to know come with a

degree of conviction lower than certainty. This does not mean that I cannot know certainties. However, there are certainties that I cannot possibly know as they resist grounding. What other properties does knowledge, as such, exhibit? For one, what I truly know is true, i.e. knowledge is *factive*. Moreover, as argued in the introduction of section 1, in accepting that knowing that here is a hand grants that we *know* the denial of the sceptical hypothesis, we can see that knowledge is *monotonic* in the sense that if p implies q, then knowing p implies knowing q. It is, as it were, closed under implication. A question arises as to how far tautologies are known. It is probably not appropriate to *claim* to know a tautology. But are they, like seemingly empirical propositions such as 'there are physical objects' entirely outside the scope of knowledge? The question is can I give proper grounds for, say, 'p implies p'? I am not sure; I am not sure what proper grounds would look like in this case. I do know what a doubt in some such cases can look like: consider the *distributive law* and it's being challenged by quantum mechanics. There is a large discussion on this topic, for example in Putnam [1968], but it is evident in the history of science that statements once thought to be necessary truths turned out to be false. (Think of the case of Euclidean geometry, also in relation to Kant's *CPR*.) We will get back to this in chapter 3.

Moving on, as we have seen, knowledge fails to be *negatively introspective*, as I can not know something, but be under the impression (like Moore) that I know it.[26]

Instances of knowledge are instances of belief. Moreover, where claiming to know is in order, i.e. where proper grounds can be given, also doubt is in order. Doubt, however, presupposes that I have a ground as well. 'I doubt p' means that I have some evidence e for $\neg p$ that I am more convicted of than p [cf. OC, 121-7]. This does not mean that I know $\neg p$. Certainty, then, is above doubt and proper grounds but this itself is a matter of circumstance. Moreover, all of these notions come in varying degrees of assurance based on their degree of conviction (beliefs), the degree of conviction of their ground (known facts and doubts), or their level of entrenchment (certainties).

[26] To completely lay out the argument that 'I don't know p implies I know I don't know p' is invalid here, consider Moore. He does not know there is a hand, let's assume. But he believes he does know. Assuming he is consistent, he does thus not believe that he does not know that there is a hand. But not believing this means he cannot know it. Therefore, he does not know it but does not know that he doesn't.

CHAPTER 2

A Formal Approach to OC

3. Probabilities, Certainties and Knowledge

INTRODUCTION. We have established towards the end of the last chapter that *belief*-like attitudes are a primitive notion in *OC*. Certainties, as well as instances of knowledge, are attitudes that come in degrees. The maximum value of these degrees is certainty.

What we aim to do in this chapter is to construct a semantic setting which lets us model (or discuss ways to model) the epistemic notions that have concerned us in the last chapter. Now, *degrees of belief* are easily associated with Bayesian epistemology and probabilities. Indeed, we want to use probabilities here to model degrees of belief or conviction. While this seems to be straightforward, we do need to say something about what we take probabilities to be, how we interpret (or analyse) them.

In a great survey of interpretations, Hájek [2012, Sec. 3] identifies three main views on probabilities:

(1) A quasi-logical concept, which is meant to measure objective evidential support relations. For example, "in light of the relevant seismological and geological data, it is *probable* that California will experience a major earthquake this decade".
(2) The concept of an agent's degree of confidence, a graded belief. For example, "I am not sure that it will rain in Canberra this week, but it *probably* will."
(3) An objective concept that applies to various systems in the world, independently of what anyone thinks. For example, "a particular radium atom will *probably* decay within 10,000 years".

The distinctions here may seem vague and the possibility of reducing one interpretation to another is not excluded. It is important to our project that we keep in mind that in taking probabilities to reflect degrees of *belief*, we inherently take them to be *subjective*.

More specifically, we will analyse probabilities (degrees of belief) in terms of betting quotients. This interpretation goes back to de Finetti [1980] and Hájek [2012, Sec. 3.3.2] sums it up as a relation between a piece of evidence E and a degree p, saying:

> Your degree of belief in E is p iff p units of utility is the price at which you would buy or sell a bet that pays 1 unit of utility if E, 0 if not E.

This analysis has consequences. Ramsey [1931, p. 15] argues that "the kind of measurement of belief with which probability is concerned is [. . .] a measurement of belief *qua* basis of action." This echoes the Wittgensteinian understanding of certainties as being reflected in our actions.[1]

Certainty is the highest among these action-guiding degrees of beliefs. The interpretation in terms of bets has further consequences for this limit degree of belief, as Leitgeb [2014, p. 2] points out:

> For example: it is morning; I rationally believe that I am going to receive an email today. However, I would not regard it as rational to buy a bet in which I would win one dollar if I am right, and in which I would lose a million dollars if I am wrong. But according to the usual interpretation of subjective probabilities in terms of betting quotients, I should be rationally disposed to accept such a bet if I believed the relevant proposition to the maximal degree of 1.

Let us look at this from a Wittgensteinian point of view. Take his example of a certainty 'cars do not grow out of the earth' and consider an agent certain of this proposition in a betting situation. Would she bet 1 dollar if this turns out to be correct against 1 million dollars if it turns out to be false? According to the status of certainty, as understood in *OC*, she would — and we could even go so far as to say that she *should*[2] — if she is to act rationally. There are two things that speak in favour of this interpretation. First, "[m]y *life* consists in my being content to accept many things" [OC, 344]. If one of these things fails, my life changes entirely. Never mind the million dollars then. Indeed, the fact that a " bet may concern an event such that, were it to occur, you would no longer value the pay-off the same way" is a consequence of this interpretation, argues Hájek [2012, Sec. 3.3.2].[3] Second, with "the word "certain" we express complete conviction, the total absence of doubt" [OC, 194] and in the *total absence* of doubting the outcome will be in her favour, it would be odd not to accept the bet.

We do aim to give a *formal* semantics here, so we need to have a clear definition of our core notion of probability. *Dutch book arguments* seem to confirm the classical axioms of probability [cf. Kolmogorov 1933] as a rationality requirement for any agent's probability distribution. The arguments show that any agent with a probability distribution not obeying the classical three axioms can be Dutch-booked, i.e. she can be put in a situation where she will certainly loose the bet [cf. Vineberg 2011]. Conversely, as

[1] There are several other consequences and issues that come with this analysis. For one, it presupposes there is precisely one price for any piece of evidence. Second, it requires the sum of the bet to be infinitely divisible. And third, we should not go so far as to take these bets literally, but rather consider them as bets we would place under a Rawlsian *veil of ignorance*, as Hájek [2012, Sec. 3.3.2] suggests: "imagine that you are to set the price for the bet, but you do not yet know which side of the bet you are to take." We will not be able to go into full detail here and refer the reader to the discussion in Hájek [*ibid.*].

[2] Recall Pritchard [2007] arguing certainties play a normative role.

[3] Hájek [*ibid.*] takes the example of a man betting 1 million dollars on the world ending at a certain date.

Kemeny [1955] has shown, if an agent obeys the axioms, she can *not* be Dutch-booked. We will thus stick with the following definition:

DEFINITION 1 (Probability Measure). *Let S be a non-empty set. An* algebra *on S is a set $\mathcal{A}$ of subsets of S s.t. $S \in \mathcal{A}$ and $\mathcal{A}$ is closed under complementation with respect to S and union.*[4] *Now let $\mu : \mathcal{A} \to [0,1]$ be a function from the algebra over S to the unit interval. We say μ is a* probability function *(or* measure*) if it obeys the following axioms:*

(P1) $\mu(X) \geq 0$, *for all* $X \in \mathcal{A}$

(P2) $\mu(S) = 1$

(P3) $\mu(X \cup Y) = \mu(X) + \mu(Y)$, *for all* $X, Y \in \mathcal{A}$ *where* $X \cap Y = \emptyset$

We call $(S, \mathcal{A}, \mu)$ a probability space.

In the sequel, we will always assume that the set S underlying a probability space is finite.

3.1. CERTAINTY AS PROBABILITY. So far we have come to an agreement on how probabilities are to be defined, how they are interpreted and how this interpretation links back to *OC*. Definition 1 is very abstract and general. We want to now take this definition and build a semantic framework in which we can consider defining the epistemic notions relevant to our project.

We should first note that we want to work with a Kripke-style semantic approach here. One reason for this is that it is well known to everyone working in (theoretical) philosophy. Moreover, knowledge, belief and even probability seem to be inherently modal notions and much work in formalizing these concepts has been done in Kripke-structures. Research in dynamic epistemic logic, for example, has produced some results significant to our goal on so-called *plausibility models*, e.g. in Aucher [2008]. If we want to adapt these results for our setting it will be helpful to work with a similar structure.

Let us begin by assuming a set Φ of atomic propositions. As is well known, we can take any such set Φ and produce a set of worlds W, the elements of which represent all the logical possibilities Φ gives rise to. (Later on, worlds may include more information than the truth values of atomic propositions and we can thus not *define* a set of worlds to be the set of (propositional) logical possibilities over a set of propositions.) For example, let $\Phi = \{p, q\}$, then we have $W = \{w_1, w_2, w_3, w_4\}$ where both p and q hold at w_1, p but not q holds at w_2, q but not p holds at w_3 and both p and q fail at w_4. We can, thus, likewise take any subset of a set of worlds W to express a proposition or *hypothesis* (we use these terms synonymously). Following our example, the hypothesis $H = P \cap Q$ saying that both p and q hold is identified with the set $\{w_1\}$, while the proposition $P = \{w_1, w_2\}$ and $Q = \{w_1, w_3\}$. Intersections can then be identified with conjunctions and complements with negations, thus giving us all the usual logical connectives as defined in terms of these two operations.

[4] A set $\mathcal{A}$ is *closed under complement* w.r.t. S if $X \in \mathcal{A}$ implies $S \setminus X \in \mathcal{A}$; it is said to be *closed under unions* if $X, Y \in \mathcal{A}$ implies $X \cup Y \in \mathcal{A}$.

Adding probabilities to the mix, our definition 1 requires an algebra $\mathcal{A}$. For any set S, its powerset $\mathcal{P}(S)$ (together with the set-forming operations) is an algebra over S, i.e. closed under complement with respect to S and union. Thus, we can take, e.g. $\mathcal{A} = \mathcal{P}(S)$ (together with the set operations) as our algebra. We will usually assume that $S = W$. In this setting, we then have a correspondence between the probability assignment of a hypothesis and a set of worlds, given as

$$\mu(H) = \sum_{w \in H} \mu(w).$$

We can now take these observations together to define what constitutes a model for the notions we are trying to define.

DEFINITION 2 (Probabilistic Epistemic Model[5]). *Given a set of propositional letters* Φ, *a* probabilistic epistemic model *is a tuple* $M = (W, \mathcal{A}, \mu, t)$ *where*

- W *is a finite set of (epistemically) possible worlds over* Φ,
- $\mathcal{A} = (\mathcal{P}(S), \cap, \cup, \setminus, \emptyset)$, *where* $S \subseteq W$,
- $\mu : \mathcal{A} \to [0, 1]$ *is a probability measure and*
- $t \in (0.5, 1]$ *is an agent's threshold of confidence*

In practice, we will often work with pointed models, (M, w), *which are pairs consisting of a model* $M = (W, \mathcal{A}, \mu, t)$ *and a world* $w \in W$ *which is the* actual world. *We will occasionally write* $w \in M$ *for* $w \in W$. *And to reiterate: unless mentioned otherwise, we take it that* $S = W$.

I take it that the first three items of the definition have been sufficiently clarified. Item four will come into play later on, but it is intuitive to assume that subjective probabilities give rise to a threshold.

On such a structure, we can properly define our notion of certainty, here understood as a monadic modal operator C on propositions, giving rise to a proposition in W.

DEFINITION 3 (Certainty[6]). *Where* $H \subseteq W$ *is a proposition,* $w \in C(H)$ *iff* $\mu(H) = 1$

We note right away that in our present setting, either $C(H) = W$ or $C(H) = \emptyset$. I.e. it is a global notion, not dependent on any world. Certainty, thus, is an all or nothing notion. This does not mean that we cannot identify different *levels of entrenchment* for certainties as we have argued for in sections 1.4 and 2.1. There are, for one, *infallible* hypothesis T (for tautology) that hold in every world, i.e. where $T = W$. It immediately follows from our definitions that $C(T)$ holds in all those cases as well. Any tautology is a certainty. This we can, if we like, identify with certainties of level 1, which are not subject to epistemic verification/falsification in any context.

A weaker class of certainties is the class of certainties that are merely in the *support* of μ, i.e. those hypothesis H where $\mu(w) > 0$ for all $w \in H$. Such a hypothesis need not necessarily hold everywhere in W.

EXAMPLE 1. Consider Ada, who's probabilities are represented in fig. 1. She is certain of every tautology. Take, for example, $p \vee \neg p$. As we can see

[5]This definition is due to Baltag [unpublished].

[6]Cf. Baltag [*ibid.*].

further, she assigns probability 1 to $p \vee q$, as the only world in which this fails is assigned probability 0. Moreover, she is certain of q, as this is in the

FIGURE 1. Entrenchment of certainties

support of her probability function, i.e. everywhere q holds, the probability is greater than 0. In terms of logical possibility, however, $p \vee \neg p$ holds in all, $p \vee q$ in three and q only in two worlds. The mere *possibility* of giving up one of these propositions, thus, ranges from virtually non-existent ($p \vee \neg p$), to quite tangible (q).

The example shows more. With any algebra, in a sense, comes a probabilistic *logic*.[7] This logic, we can take to be the subjective logic of the agent.

Recall that the set underlying the algebra is certain to the agent and that certainties form the background against which true and false are distinguished.

> But I did not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false. [OC, 94]

We can now let the logic differ radically from classical logic by allowing that the underlying set $S \neq W$. Consider the subjective logic of an agent as the logic of her language-game. Propositions can in this way be expunged from the language-game: "we expunge the sentences that don't get us any further" [OC, 33]. Now if we take the set S to be a proper subset of W, then tautologies are expunged from the language. We have agreed that a hypothesis' probability is equal to the sum of the probability of the worlds at which it holds, but if one of these worlds is undefined, which is the case when the world is in $W \setminus S$, then the hypothesis' probability is undefined as well. The notion of certainty will thus no longer be global, but only hold with respect to a subset $S \subset W$. A tautology holding at every world, thus, will have no probability as soon as $S \subset W$. Of course, we can easily recapture certainty of tautologies by requiring that for all hypotheses H, $\mu^*(H)$ (a new measure) be equal to the sum of all worlds $w \in H$ where $w \in S$. We might, however, want to get rid of tautologies seeing as it is generally assumed they don't get us any further. In the prequel we will assume $S = W$ unless stated otherwise.

In our example above, Ada's logic exhibits some strange properties if we expunge $\neg p \wedge \neg q$ from her language. For example, in Ada's algebra, the complement of $p \vee q$ will be the empty set. This means, its negation is *considered* a contradiction, which puts it on one level with a tautology

[7] Probabilistic logic is usually defined over a language directly [for a survey cf. Demey *et al.* 2014], but what we are considering here is not a logic, *per se*, but a subjective *method of inquiry*.

from Ada's point of view. On the other hand, q is just as certain, but still subjectively distinguishable from a tautology. $p \vee q$ does not have the form of a tautology, so if we take Φ to include only empirical propositions, we *can* say "propositions of the form of empirical propositions, and not only propositions of logic, form the foundation of all operating with thoughts (with language)" [OC, 401]. Notice moreover that we have thus allowed for another class of certainties: the propositions equivalent to S do not hold everywhere in W but are also true at worlds outside the support of μ, i.e. where $\mu(w) = 0$.

3.2. DEGREES OF BELIEF AND DOUBT. We now turn to the lesser degrees of belief. We have already introduced a threshold t of confidence in the previous section and we will make use of this now. Following Baltag [unpublished], we let the threshold $t \in (0.5, 1]$ and we will say that an agent is *highly confident* in a hypothesis H iff $\mu(H) \geq t$. Now, ""I believe" is an 'expression"' [OC, 180] that "has subjective truth" [OC, 179] and comes in degrees of assurance:

> I make assertions about reality, assertions which have different degrees of assurance. How does the degree of assurance come out? What consequences has it? [OC, 66]

Taking probabilities to reflect degrees of *belief*, we have to admit that we have not fully captured what belief is, as beliefs, *per se*, require *assurance*. Giving assurance can be taken as "giving a degree of certainty" [OC, 386] and implying the possibility of giving evidence.

Now, given a hypothesis H which comes with some degree of belief, what role does an incoming piece of evidence play? We take the standard approach of taking *conditional probability* as modelling the entry of a new piece of evidence. That is, given the probability $\mu(H)$ an agent assigns to a hypothesis H and the probability $\mu(E)$ she gives to a piece of evidence E, we can calculate the probability $\mu(H|E)$ of H *given* E as

$$\mu(H|E) = \frac{\mu(H \cap E)}{\mu(E)},$$

provided, of course, that $\mu(E) > 0$. With this notion of conditional probability in place, Baltag [unpublished] and Baltag and Smets [2008b, pp. 9–10] define a (least) *degree of safety* an agent assigns to a hypothesis at a state.

DEFINITION 4 (Degrees of Safety). *The* degree of safety $d_w(H)$ *an agent assigns a hypothesis H at world w is defined as*

$$d_w(H) := \min_{w \in E \subseteq W} \mu(H|E)$$

These degrees of safety can be identified with degrees of assurance or degrees of certainty. They are not a mere degree of belief, but rather reflect an agent's readiness to give up a belief; the higher the degree of safety, the lower the chance an agent will give up her belief in light of new evidence (that is true at a state).[8] This notion is *local*, i.e. dependent on a world, as

[8]The attentive reader might, at this point, be inclined to ask for a stronger, *absolute* notion of certainty which holds for a hypothesis H iff $\mu(H|E) = 1$ for all pieces of evidence where $\mu(E) > 0$. However, as Baltag [unpublished] assures us, in our classical probability

opposed to the global notion of degrees of belief. Note, however, that at the world in question, the piece of evidence defining the degree of safety has to be true; thus, degrees of safety reflect a degree of *truth* at a world, which depends on the agent's subjective probability distribution. Extending this, we can, following Baltag and Smets [2008b, pp. 9–10], define, for any $t \in [0,1]$, a modality $\Box^t$ reflecting the degree of safety as

$$w \in \Box^t(H) \quad \text{iff} \quad d_w(H) > t.$$

Beliefs, understood as giving degrees of assurance, can then be taken to be a range of modalities $\Box^t$, where $t \in (0.5, 1]$. Thus let $w \in B(H)$ iff there is $t \in (0.5, 1]$ such that $w \in \Box^t(H)$.

Baltag [unpublished] provides us with a way to calculate degrees of safety. Note first that for any H with $\mu(H) = 1$, $d_w(H) = 1$ as certainty is absolute. Now, if $\mu(H) < 1$, then we have two cases. If H is false at s, then $s \in W \setminus H$, so $\mu(H|W \setminus H) = 0$ and hence $d_w(H) = 0$. Now, if $s \in H$, then consider the set $E = \{w\} \cup W \setminus H$ which is clearly the strongest proposition that speaks against H that is true at w. Then

$$d_w(H) = \mu(H|E) = \frac{\mu(H \cap E)}{\mu(E)} = \frac{\mu(w)}{\mu(w) + 1 - \mu(H)}$$

It should be noted that thus we cannot construct the so-called *lottery paradox* [as described in van Fraassen 1995] for degrees of safety. Assume a fair lottery with n tickets. Consider a set of worlds $W = \{w_1, \ldots, w_n\}$ s.t. ticket i wins at w_i (and all others loose there). Say Ada assigns equal probability to each ticket being drawn, i.e. for all tickets i, $\mu(i) = \frac{1}{n}$. Thus, she assigns probability $\frac{n-1}{n}$ to any ticket *not* winning. Then for w_i, $d_{w_i}(\neg i) = 0$. For all other worlds $w_j \neq w_i$,

$$d_{w_j}(\neg i) = \frac{\mu(w_j)}{\mu(w_j) + 1 - \mu(\neg i)} = \frac{\frac{1}{n}}{\frac{1}{n} + 1 - \frac{n-1}{n}} = \frac{\frac{1}{n}}{\frac{1}{n} + \frac{1}{n}} = \frac{1}{2}$$

Whichever world is the actual world, given a reasonable threshold of $t \geq \frac{1}{2}$, Ada will not believe any ticket to lose or win. Still, she is highly confident that any given ticket, say her own, is going to lose. Note, moreover, that when take the intersection

$$\bigcap_{1 \leq i \leq n} \Box^{0.5}(\neg i)$$

---

setting — accepting axioms (P1)–(P3) (cf. p. 31) — this is equivalent to certainty understood as probability 1.

PROOF. If $\mu(H|E) = 1$ holds then in particular $\mu(H|W) = 1$ and since $\mu(W) = 1$ and $H \cap W = H$, also $\mu(H) = 1$. The same holds for any subset $S \subseteq W$ for which μ is defined. To see the converse, let $\mu(H) = 1$ and let $E \subseteq W$ be such that $\mu(E) > 0$. Let $A^+ := \{s \in A \mid \mu(w) > 0\}$, then $\mu(A) = \sum_{w \in A} \mu(w) = \mu(A^+)$. Then $\mu(E \cap H) = \mu(E)$, as for any $w \in E^+$, $s \in H^+$, for otherwise $\mu(\{s\} \cup H) > 1$, which is forbidden by definition. Hence, $E^+ \subseteq H^+$, thus $E^+ \cap H^+ = E^+$. But $E^+ \cap H^+ = (E \cap H)^+$, since $w \in E^+ \cap H^+$ iff $\mu(w) > 0$ and $w \in E$ and $w \in H$ iff $w \in E \cap H$ and $\mu(w) > 0$ iff $w \in (E \cap H)^+$. Thus, $(E \cap H)^+ = E^+$ and therefore $\mu(E \cap H) = \mu(E)$. But then $\mu(H|E) = \frac{\mu(E \cap H)}{\mu(E)} = \frac{\mu(E)}{\mu(E)} = 1$. □

we will get an empty set. Observe further that $H \subseteq G$ implies that $B(H) \subseteq B(G)$: Let $H \subseteq G$ and $w \in B(H)$, then $d_w(H) \geq t$. Since $H \subseteq G$, $\mu(H) \leq \mu(G)$ (assuming μ is defined on all of W). Take any piece of evidence E true at w, then $\mu(G|E) \geq \mu(H|E) \geq t$. Thus, $w \in B(G)$. This shows that B is a normal modal operator (if μ is defined on all of W) and, thus (assuming the rule of necessitation holds), closed under conjunction [cf. Blackburn *et al.* 2001, p. 35]. Note that we have the intuitively correct conclusion that an agent believes some ticket will win, no one ticket will be the winner, but her beliefs are closed under conjunction. This fails for notions of belief that are identified with a mere probability assignment above the threshold.

Let us now turn to doubt. When discussing the betting interpretation of probability on page 30, we gave two arguments for its correctness as an interpretation of certainty in *OC*. (i) That "my *life* consists in my being content to accept many things" [OC, 344] suggests that certainties are a way of looking at the world which would change entirely if we began to *doubt* a certainty. (ii) That with "the word "certain" we express [...] the total absence of doubt" [OC, 194] suggests that at times doubt cannot even apply. Just as we had two notions of belief — a mere attitude we identified with probabilities and a grounded attitude we identified with degrees of safety — this seems to suggest there are two ways in which doubt can be absent. First, the absence of doubt can be due to the strongest from of evidential support (the way I understand the world); second, it can be due to the concept of doubt not applying in a given situation (the total absence of doubt).

"Doubt comes *after* belief" [OC, 160] and requires *grounds* [cf. OC, 4], i.e. evidence to the contrary. Thus, if we want to say that an agent doubts a hypothesis H, she should (at least) *believe* she has some evidence E, given which her degree of belief in H is 0 (or less than t). To tie this to our two notions of absence of doubt, we have to take it that if the evidence contradicting H has probability 1, doubt is no longer present. And in the limit case, where a hypothesis is certain, the *concept* of doubt no longer applies (to its negation):

> [...] Or are we to say that certainty is merely a constructed point to which some things approximate more, some less closely? No. Doubt gradually loses its sense. This language-game just *is* like that. [OC, 56]

Thus, to doubt some hypothesis H we need $\neg H$ to not be certain and some piece of evidence E that is less than certain, but some form of grounded belief in E such that H given E has probability 0 (or is less than t). Formally:

DEFINITION 5 (Doubt). *Let $H, E \in \mathcal{A}$. H is* doubted *at w, i.e. $w \in D(H)$ iff*

$$\mu(\neg H) \neq 1 \wedge \exists E \subseteq W(\mu(E) < 1 \wedge d_w(E) > t \ \wedge \ \mu(H|E) = 0)$$

Likewise, H is weakly doubted *at w, i.e. $w \in wD(H)$ iff*

$$\mu(\neg H) \neq 1 \wedge \exists E \subseteq W(\mu(E) < 1 \wedge d_w(E) > t \ \wedge \ \mu(H|E) < t)$$

With this definition in place, we immediately "exempt certain propositions form doubt" [OC, 88, cf. 341], viz. the certainties. After all, certainty implies absolute certainty (cf. footnote 8), thus for any piece of evidence

E (with positive probability, as required for doubt) and any certainty C, $\mu(C|E) = 1$. Moreover, negations of certainties cannot be doubted and certainty cannot ground a doubt: "This body of knowledge has been handed on to me and I have no grounds for doubting it, but, on the contrary, all sorts of confirmation" [OC, 288].

Scepticism, too, is ruled out. Already axiom (P2) requires that a rational agent must assign probability 1 to the set S underlying her algebra. Moreover, S may not be empty and, thus, there has to be some non-contradictory proposition the agent is certain of. Therefore, there are at least two propositions that are immediately exempt from doubt for any rational agent: the underlying set S and its complement $\neg S$, simply due to the fact that S is certain. Hence, radical scepticism, i.e. doubting *everything* (to some degree at least),[9] is entirely ruled out by our requirement for rationality.

We have previously allowed to take a proper subset of W to be the set underlying the algebra, i.e. to not assign any probability to some world. We have taken this as an agent expunging certain propositions from the language-game. For the same reason as above, we cannot expunge all propositions.

3.3. KNOWLEDGE. Scepticism is entirely ruled out by our system, which was deemed rational by Dutch-book arguments. Strangely, *knowledge* did not play a part in this observation. As we saw, there is a class of certainties that hold everywhere on the agent's probability space, yet not everywhere on the set of worlds. This class is empty for agents who expunge no propositions from their language. This sort of scepticism (although 'scepticism' may not be the correct term here), as opposed to radical scepticism, makes sense, can be warranted and can turn out to be false.

This is where the objective dimension enters the picture and knowledge, as we read, can be established only on objective criteria [cf. OC, 15, 16, 179, 245]. Taking an external point of view, we can single out one of the logical possibilities $w_a \in W$ which we call the *actual world*. Whatever holds there, holds *in fact*. Clearly, being judged on the actual world, knowledge must be a local notion but certainty does not qualify as a good definition. As we have seen in sections 1.4 and 2.1, certainties can sometimes be instances of knowledge ('My name is M.B.'), sometimes not ('There are physical objects'). Clearly, certainties for one agent may be knowledge for another (you know my name is M.B.); the fact that 'my name is M.B.' is something *I* know is not directly mentioned in *OC*, but following Williams [2007] in subsection 1.1 we have established that sometimes is can be, or has to be intuitively. Moreover, less than certain propositions can be known ('Cows have four stomach compartments') and even *said* to be known. Hence, we want our notion of knowledge to neither imply, nor be implied by certainty. It should be noted that the following discussion depends on our interpretation (of certainties being sometimes known) forcing a strong connection between knowledge and certainty.

[9] We can, similar to our degrees of belief, define degrees of weak doubt depending on the threshold t.

Surely, some vote of confidence in a proposition needs to be presupposed in order to say someone *knows* it. After all, "[w]hat I know, I believe" [OC, 177]. Requiring further that knowledge be *factive*, i.e. can only apply to propositions that hold at the actual world, a straightforward way to define knowledge would be to take it as *true high confidence*, following Baltag [unpublished]. That this will not do can be shown with the following Gettier-type example.[10]

EXAMPLE 2. Say the feeling creeps up on Ada that maybe someone *could* (at some point in time) go to the moon. Living in Wittgenstein's times where this was not a common-place belief (let's assume), she is *certain* that this feeling can only be due to the fact that there is an actual possibility of going to the moon or her being insane ($\mu(m \vee i) = 1$). She is highly confident that she is not insane, say $\mu(\neg i) = 0.9$ where her threshold $t = 0.9$. Based on this, she should be highly confident that she is a sane person who believes in space travel ($\mu(m \wedge \neg i) = 0.9$). The model in fig. 2 reflects the situation,

FIGURE 2. Space travel or insantiy?

with the actual situation highlighted by double boxing. In reality, Ada is *both* right about space travel, but sadly also insane. She is certainly highly confident in the possibility of space travel and it is, in fact, true that space travel is possible as we now know. Can she, thus, be said to *know* that space travel is possible?

No, would be Wittgenstein's answer, I believe. If pressed to justify or give grounds for her supposed knowledge, she would have to rely upon the fact that she is not insane, which is less certain than her belief in the possibility of space travel. After all, $\mu(\neg i) = 0.9$ and $\mu(m) = 0.91$. Even worse, that she is insane could, in our example, be empirically established by, say, an EEG. Then, given we can use conditional probabilities to model this learning, the resulting situation would be captured in fig. 3. Here we

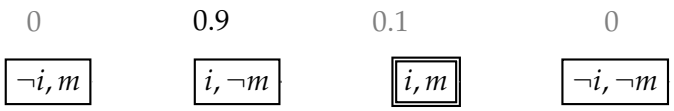


FIGURE 3. Learning you are insane.

can see what Ada's epistemic state would look like, had she *made a mistake*. She has become highly confident it is impossible to go to the moon.

[10]Famously, Gettier [1963] showed with his examples that knowledge is not justified true belief. The example below is adapted from Baltag [unpublished].

What is missing in the notion of true high confidence is, as was the case with belief, an assurance, or the possibility to give *evidence*. More precisely, we need a *good* notion of evidence or grounding, as we saw the last one fail. In particular, we do not want *true* information to be able to defeat our knowledge; "it needs to be *shewn* that no mistake was possible" [OC, 15], and if, in showing that I did not make a mistake true evidence to the contrary pops up, it will not count as having been shewn. We begin by considering a world w and any piece of evidence E that holds at w, i.e. $w \in E$. No such piece of evidence should lead to a loss in high confidence in a hypothesis H, if we want to exclude the possibility of being mistaken about H. In formal terms, $\mu(H|E) \geq t$. Following Baltag [unpublished], we can define an operator U that expresses that a given hypothesis is undefeated.

DEFINITION 6 (Undefeated Proposition). *Let $H \in \mathcal{A}$, then*

$$w \in U(H) \quad \textit{iff} \quad \forall E \in \mathcal{A}((\mu(E) > 0 \ \wedge \ w \in E) \textit{ implies } \mu(H|E) \geq t)$$

If a hypothesis is undefeated, mistakes are *logically* excluded [cf. OC, 194]. Moreover, undefeatedness is just a special case of our degrees of safety, so undefeated propositions are believed. One would think this makes a good candidate for knowledge, but it turns out that false certainties are, because absolutely certain, instances of false undefeated propositions. In fact, as Baltag [unpublished] points out, falsely believed certainties are the only propositions that are false and undefeated. We could say, undefeatedness is a *quasi-factive* notion, as every hypothesis that is less than certain and undefeated is guaranteed to be true.

CLAIM. *If $\mu(H) \neq 1$, then $U(H) \rightarrow H$.*[11]

PROOF. Suppose $\mu(H) < 1$ and some threshold $t \in (0.5, 1]$. Since $\mu(H) < 1$, there is a set $E = S \setminus H$ s.t. $\mu(E) > 0$. Suppose for a contrapositive that $w \notin H$ then $w \in E$. Now

$$\mu(H|E) = \frac{\mu(H \cap E)}{\mu(E)} = \frac{\mu(\emptyset)}{\mu(E)} = 0 < t$$

Therefore, $w \notin U(H)$. □

Thus, testing for possible mistakes is not a method of verifying certainties, but it is for all less than certain hypotheses. Nevertheless, undefeatedness is not a candidate for knowledge, as we have argued that certainties (at least in certain circumstances) can be known. (Whether this knowledge can be coherently expressed is a different matter!) Now it would be premature to conclude that knowledge is simply *true* undefeatedness. While it is easy enough to define an operator TU, where $TU(H)$ holds at w iff $w \in U(H)$ and $w \in H$, such an operator has rather strange properties. Clearly, it is factive, but it is not *normal*, i.e. not closed under conjunction.

CLAIM. $TU(A) \cap TU(B) \not\subseteq TU(A \cap B)$[12]

[11] Cf. Baltag [*ibid.*].

[12] Cf. Baltag [*ibid.*] for a counter example.

Now, perhaps one might be willing to accept that knowledge is not closed under conjunction. To my knowledge, however, there is nothing in *OC* that speaks for or against closure under conjunction. I believe this closure to be highly intuitive and, thus, would require a strong argument for a denial of it in *OC*. Here, we simply take it as given that knowledge is closed under conjunctions.[13]

What we can do, following Baltag [*ibid.*], is add a notion of *justification*, or proper ground in Wittgenstein's terms. We say that J is a justification for H iff $J \to H$ (or $J \subseteq H$). If we now require that in order to know H, H must be true and undefeated and have a true and undefeated justification, we won't get far. Formally, this would mean

$$w \in H \cap U(H) \wedge \exists J \subseteq W.(J \wedge U(J) \wedge (J \to H))$$

which, since every proposition could justify itself, gives us little more than truth plus undefeatedness. We could, in the spirit of Wittgenstein, require the justification to be more certain than the hypothesis, i.e. define

$$w \in K(H) \quad \text{iff} \quad w \in H \cap U(H) \wedge \exists J \subseteq W.(\mu(J) > \mu(H) \wedge J \wedge U(J) \wedge (J \to H))$$

Now, we would loose all knowledge of certainties, as there can be no justification for them that is more certain. If we adopt a non-strict version of the above, we again allow every proposition to justify itself.

We are in need of a stronger notion of justification. To this aim, let's say a proposition is *resilient* when it is undefeated whenever it is true, i.e.

$$w \in R(H) \quad \text{iff} \quad w \in H \subseteq U(H)$$

We may then say that knowledge requires a true and resilient justification.

DEFINITION 7 (Knowledge).

$$w \in K(H) \quad \textit{iff} \quad \exists J \in \mathcal{A}.(w \in J \subseteq H \textit{ and } J \subseteq U(J))$$

Again, we may want to require that J be more certain than H but this would also exclude certainties from knowledge.[14] The notion we have defined here does indeed rule out possible mistakes. Any instance of knowledge now requires a justification that is true, resilient, and hence undefeated. Moreover, it is equivalent to repeating the possible mistakes-test and infinite number of times on a true proposition.[15]

We need to stress, however, that the notion of justification we have here does not exhaust what Wittgenstein means by 'proper grounds'. For example, recall our discussion on certainties as a foundation for knowledge in subsection 1.1. There, we have seen that sometimes *actions* can be a form of justification.

---

[13] As Baltag [unpublished] points out, another problem with the TU operator is that it is not positively introspective. This, too, is rather counter-intuitive.

[14] Perhaps we can think of just requiring J to be *at least as certain* as H, but Wittgenstein seems to be against justification that is "no *more* certain" [OC, 1, my emphasis].

[15] Baltag [unpublished] assures us that the definition of K above is equivalent to defining it in terms of the following, infinite conjunction

$$K(H) := H \cap U(H) \cap U(U(H)) \cap U(U(U(H))) \cap \ldots$$

which expresses that H is true, there was no mistake made in believing H, there was no mistake in believing there was no mistake in believing H, and so on.

> What *counts* as its test? "But is this an adequate test? And, if so, must it not be recognizable as such in logic?"—As if giving grounds did not come to an end sometime. But the end is not an ungrounded presupposition: it is an ungrounded way of acting. [OC, 110]

To be sure, degrees of belief, i.e. probabilities, are here taken to be the *basis* of action, but are not themselves to be confused with actions. In order to incorporate actions as justifications into a definition of knowledge, we need a firm grasp on events, actions, utterances, and such. We will get back to this in subsection 4.3, where we discuss an approach that touches on this subject.

It should be duly noted that our definition 7 gives us a notion of knowledge that is defeasible [cf. Baltag unpublished]. I.e. it does not satisfy negative introspection 'if I do not know p then I know I do not know p.' This is important. Wittgenstein [OC, 179] claims "[i]t would be correct to say: "I believe' has subjective truth". Not only does this mean that 'I believe p' gives us a degree of truth of p for an agent, but it seems to the agent to be true what she believes, i.e. p. Thus, she will believe to know, since she believes p and that she bears a distinct relation to p, *viz.* that of being acquainted with it as a fact (cf. subsection 2.3). This seems to suggest that we have the following validity

$$B(H) \subseteq BK(H)$$

As we have seen in footnote 26, if we had negative introspection, we would get that

$$BK(H) \subseteq K(H)$$

We can see that, thus, negative introspection would imply that believing H implies knowing H. This is known in the literature as the Perfect Believer [cf. Baltag and Smets 2008a, p. 31] and it is not very intuitive. Thus, it is good to follow Wittgenstein [OC, 12] and not to forget "the expression "I thought I knew"."

We want to conclude this chapter with some examples, taken from Baltag [unpublished], showing that our definition 7 of knowledge does not imply and is not implied by certainty, and that knowledge is not negatively introspective, as required in section 2.4.

EXAMPLE 3. Consider another insane space travel-example where probabilities are distributed as in fig. 4. The real world w_a, again, is the i, m world. In this model, Ada *knows* i, as i justifies itself,[16] is true (at w_a) and

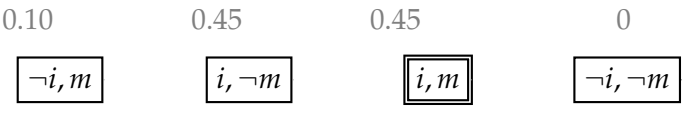


FIGURE 4. Knowledge without certainty.

[16] Note that not any given proposition can justify itself, only true undefeated propositions can.

undefeated (if we let $t = 0.8$). Her degree of certainty in i is only 0.9, so she is not certain of i. In fig. 5, Ada *is* certain of $\neg m$ but does not, in fact, know it as it is false.

FIGURE 5. Certainty without knowledge.

She does still *believe* she knows $\neg m$! For consider the set $K(\neg m)$. $\neg m$ can, again, justify itself and is undefeated because it is certain. So $K(\neg m)$ is simply the set of all states where $\neg m$ holds. The proposition these states can be identified with, are certain and thus believed. So Ada *believes she knows* $\neg m$, but does not *know* $\neg m$, which provides a counter example to negative introspection which states

$$\neg K(H) \subseteq K\neg K(H).$$

We can see this isn't valid here as she does not know $\neg m$ (because it's false) but she does not know that she doesn't know $\neg m$. This is because she does not believe to *not know* $\neg m$ and hence (since knowledge requires belief) she cannot know that she doesn't know $\neg m$.

4. Multi-Agent Systems

Throughout section 3, we have looked at the possibility of modelling one person's certainties, beliefs, doubts etc in the spirit of *OC*. We could have considered this as a subjective perspective, so long as we avoid calling one world the actual world. Adding the actual world makes the models from the last section objective.

However, there are passages in *OC* that suggest that a static, objective, one-agent perspective is not enough. For one, it is of interest to Wittgenstein how different people's beliefs interact, i.e. how the epistemic notions hold up in a multi-agent setting. For example, the story of Moore and the king [cf. OC, 92] seems to be of importance as the subjective approach of *OC* allows for disagreement at the level of certainties. Moreover, it seems most examples of certainties are given from a first-person point of view, which cannot, as we will shortly see, be modelled as a multi-agent extension where we avoid singling out a real world as was the case in the single agent setting. Last but not least, *OC* is concerned with knowledge *claims*, which are to be taken as actions people can perform, and allows for shifts in certainties. Thus, it will be interesting for us to look at the prospects of adding dynamics to our setting.

4.1. MULTI-AGENT EXTENSIONS AND DISAGREEMENT. Before we turn to a subjective or dynamic perspective, we want to allow for more agents, so as to also allow for higher order knowledge (i.e. knowledge of knowledge) of others. This will help with allowing for third-person knowledge-claims, as discussed in chapter 1.

Clearly, we could just take a set A of agents and require every agent to have a probability distribution over W (or a subset thereof). However, this would not be very realistic. If every agent is assigned one probability distribution, we can either allow for an agent to be aware of another agent's distribution, or not. In the first case, it bars us from modelling agents being wrong about other agent's beliefs. In the second case, nothing is gained in terms of epistemic interactions. It seems appropriate to avoid both cases.

More appropriate and more expressive ways to model multi agent probabilistic epistemic logic can be found in Kooi [2003] and van Benthem *et al.* [2009]. Both have models with probabilistic information but, as opposed to us, they also have an epistemic accessibility relation to model knowledge and belief. We will just leave this out of all subsequent definitions and concentrate wholly on the probabilistic parts. For the probabilistic information, van Benthem *et al.* [2009, p. 69] require every agent to be assigned a probability distribution over W at every world in W. This immediately lets us model agents being wrong about other agent's beliefs (but also about their own beliefs). As a consequence, as we mentioned in subsection 3.1, worlds now contain more information that mere truth-tables, as they contain probabilistic information as well. Kooi [2003, p. 286] has a looser definition, in that he allows for the probability distribution of an agent at a world to be a partial function from W to the unit interval. I.e. at some worlds an agent may not assign any probability to some world (recall here the case in subsection 3.1 where S was a proper subset of W). Kooi [*ibid.*, p. 387] argues that with this partiality "one can have *ignorance about probabilities*, not just *ignorance in terms of probabilities*." We have interpreted this differently and could say an agent at a world may *expunge* certain propositions from her language. We also want to go one step further in terms of generality and loosen the requirement that every agent be assigned a probability distribution at *every* world in W. In practice, it will be enough to require a probability distribution at the actual world in most cases. We thus propose the following definition, which is general enough to include the definitions of Kooi [2003] and van Benthem *et al.* [2009].

DEFINITION 8 (Multi-Agent Model). *Given a set A of agents and set of propositions Φ, a* multi-agent probabilistic epistemic model *is a tuple $M = (W, \mu, t, V)$ where*

- *W is a (finite) set of (epistemically) possible worlds,*
- *$\mu : W \to A \rightharpoonup (W \rightharpoonup [0,1])$ assigns a (possibly partial) probability measure over (parts of) W to every agent at every world,*
- *$t : A \to (0.5, 1]$ assigns each agent a threshold.*
- *$V : \Phi \to W \to \{0,1\}$ is a valuation.*

At times, we will take pointed *multi-agent models, by which we mean a pair (M, w_a) with $w_a \in M$. We will write $\mu_a(w)$ for the probability function of agent a at world w and t_a for $t(a)$. Moreover we assume that $\mu_a(w)$ is defined for all agents at the actual world.*

REMARK 1. This differs a lot from the models of definition 2. First, an agent's probability distribution is no longer a uniform measure over the set of worlds. Instead, we have probability distributions for agents that

depend on a world. Certainty, thus, is no longer a global property in the models and will have to be evaluated on an actual world, unless we make some further assumptions about μ. Pairs of agents and worlds now make up an agent's algebra. This is also the reason we require a valuation function V here. The set of worlds W is, as mentioned, no longer exhausted by the logical possibilities of propositional logic. We can have two worlds w, v satisfying exactly the same atomic propositions, yet still containing different information about an agent a's probability distributions $\mu_a(w), \mu_a(v)$ at those worlds. We then take it that for a proposition $p \in \Phi$, $p \in w$ iff $V(p)(w) = 1$; the other connectives are then again defined in terms of complement and intersection. Hypotheses, as before, then have a probability equal to the sum of the probabilities of the worlds in which they hold. Moreover, the $\rightharpoonup$ sign is important here. As discussed, we do not necessarily require every agent to be assigned a probability measure at every world. We also do not necessarily require every probability measure of an agent at a world to assign a probability to every state in W, as we already did in subsection 3.1. Another consequence of this is that agent's may have no introspection with respect to their own probabilities, i.e. they are not necessarily aware of what their measures look like.

EXAMPLE 4. Let's see how our proposed system works. We can recapture the examples above. We simply let the old probability distribution over an algebra be the distribution at the actual world, as in fig. 6. There, an arrow labelled '$a : 0.9$' pointing from w_3 to w_2 is to mean 'agent a assigns probability 0.9 to w_2 at w_3'. Ada is agent a and μ_a is undefined for all worlds other than w_3. We represent the probability distribution $\mu_a(w)$ for an agent at a world with arrows going out from w and labelled with the probability a assigns to respective states. For presentation purposes, we omit arrows for 0 probability states.

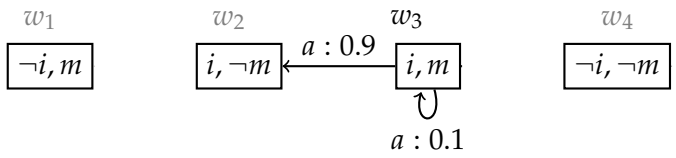


FIGURE 6. Insanity and space travel in a multi-agent setting.

These models can be more informative. For example, we can add an agent b (Bob) who is certain Ada is not insane and indifferent to space travel, as seen in fig. 7.

Rearranging things a little in fig. 8, we can see the support sets, i.e. the set of states an agent assigns a probability > 0, of the two agents are disjoint. This, again, speaks for a possible worlds semantics, as this shows precisely that disagreement about *one* certainty (in this case i) can lead to wholly disjoint support sets. Now, Bob does not *doubt* Ada is insane, as he is certain she is not. They simply deeply disagree on the matter of Ada's sanity.

Now suppose that b fully trusts a, say because he thinks a agrees with him as in fig. 9 or simply because he thinks she's sane. He takes what a

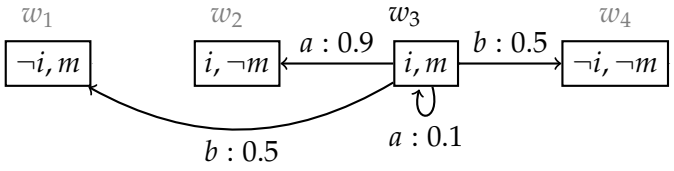


FIGURE 7. Ada and Bob.

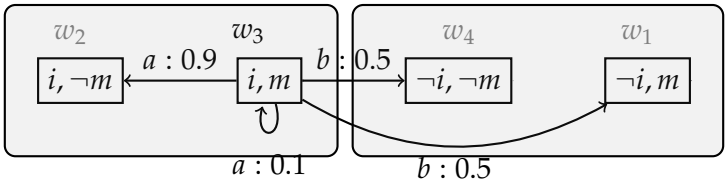


FIGURE 8. Disagreement.

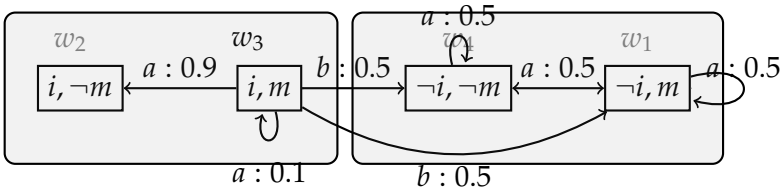


FIGURE 9. Disagreement and wrong beliefs.

says for granted as a fact. We can model such learning, as before, with conditional probabilities. Say that a informs b that i is the case, which in this case is also true. b's actual probability $\mu_b(w_3)$ will now change to a new probability $\mu'_b(w_3)$ and we can calculate this change for every world. All worlds will, however, be undefined as for any world w_i we get

$$\mu'_b(w_3)(w_i) = \mu_b(w_3)(w_i|i) = \frac{\mu_b(w_3)(w_i \cap i)}{\mu_b(w_3)(i)}$$

and $\mu_b(w_3)(i)$ is undefined. Hence, if we assume that every agent needs to have an actual probability distribution at the actual world, b we have to concede that we cannot decide (for b) what his new probability distribution looks like. From a technical point of view, then, by requiring that b have a probability distribution at w_3, the models seen in fig. 10 are, technically, equally good outcomes of the utterance of i. However, in fig. 10(a), b comes to agree with a while in fig. 10(b) he does not change his beliefs at all.

The above examples are important. For one, they reinforce our choice of Kripke models. It is inherent in them that if we have two sets of worlds, one of which is supposed to contain the negation of a formula valid in the other, these two sets will be disjoint. The sets of formulas valid in these sets are, however, not disjoint, as tautologies hold at all worlds and are thus

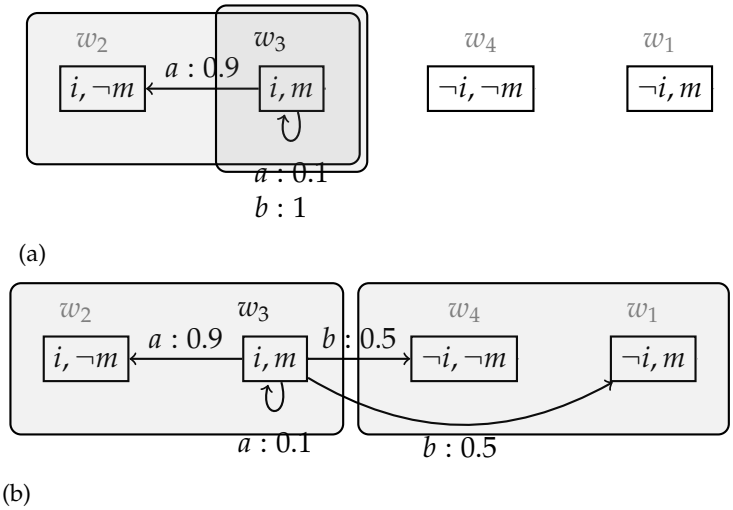


FIGURE 10. Equally good outcomes.

common to both sets.[17] This should be strongly reminiscent of *OC*, as the disagreement, here, is on *empirical* matters. The propositions $p \in \Phi$, after all, are not logical so they can be taken as empirical propositions.

Another thing the above example should echo is the story of Moore and the king [cf. OC, 91–93]. The lack of any procedure to determine what b will do if he is told i reflects the indeterminacy Wittgenstein is getting at in those passages. It is not determined by anything in our system what b will do and he might, in fact, even be seen as exiting the picture, the language-game altogether. That is not to say that b himself cannot determine what he wants to do (i.e. he may choose to believe a, or choose to stick with his distribution, or choose to change his distribution in any way he wants), but the outcome of the disagreement between a and b is not predetermined by our framework.[18]

Third, not knowing which world is the actual world, any agent has to concede that her view is just as good as the opposite (although only one is correct) and there is no way to put one view over the next without reference to the real world. With a real world in place, however, we can say that some probability distributions are superior to others, ranging from absolute superiority (assigning $\mu_a(w_a)(w_a) = 1$) to absolute inferiority (assigning $\mu_a(w_a)(w_b) = 1$ where w_b is such that for any $p \in w_a$, $\neg p \in w_b$ and for all $\neg p \in w_a$, $p \in w_b$). We have argued towards the end of subsection 1.1 that superior systems of certainty, depending on accordance with the real world, are possible [cf. also OC, 286].

4.2. A FIRST-PERSON PERSPECTIVE. In much of what we have done so far, we take an external perspective. That is, we model epistemic states

[17] Whether these tautologies figure into the logic of the language of the agents is, as seen in subsection 3.1, dependent on how we want to treat them.

[18] Further work on persuasion-tactics will have to be done in order to determine an outcome and for plausibility models this has been done in, e.g. Baltag and Smets [2009].

of agents from a god-like perspective. Now, *OC* is not written in a way that would suggest such a perspective. Rather, Wittgenstein takes into consideration a number of subjective perspectives. The theory he expresses is built from these subjective perspectives and so our models should be built from subjective perspectives as well. Aucher [2008, p. 16] reminds us that "the models built will be quite different whether the modeller is one of the agents [$a \in A$] or not." Luckily, Aucher [*ibid.*][19] gives us a way to construct such models built by one of the agents.[20] First, however, let us see what points of view there are. Aucher [2008, pp. 16–17] identifies three:

(1) The *internal* point of view, where the modeller is one of the agents.
(2) The *external* point of view, where the modeller is not one of the agents.
    (a) The *perfect* external view, where the agent has perfect knowledge of the situation.
    (b) The *imperfect* external view, where the agent does not have perfect knowledge of the situation.

Aucher [*ibid.*, p. 17] argues that "[b]ecause we proceeded by successive dichotomies, we claim that the internal, the perfect external and the imperfect external points of view are the only possible points of view". He reminds us, too, that if we consider only one agent, the perfect external point of view collapses into the internal point of view, provided that we do not make reference to the actual world. Thus, the models of sections 3.1 and 3.2 can be considered (imperfect) internal models, the models of section 3.3 can not, as they make reference to the real world in order to give knowledge the objective dimension it requires. The models of section 4.1 are wholly external.

The interesting case for us, then, is the internal point of view in the presence of multiple agents. A widely accepted internal approach is the AGM belief revision theory,[21] which cannot, however, account for beliefs about other's beliefs. Aucher [2008, p. 18] aims to generalize the AGM approach, which focuses on possible worlds, by extending the notion of a possible world to that of a *multi agent possible world*, which then also includes information about other agent's beliefs.

First, Aucher [*ibid.*] singles out an agent $y \in A$ (for *y*ou). A multi agent possible world is defined as a pair (M, w) consisting of a finite pointed (standard) Kripke-model $M = (W, R, V, w)$ *generated by* w, together with the requirement that R_a be serial, transitive and euclidean for all $a \in A$ and $R_y(w) = \{w\}$ and there is no v and $a \neq y$ such that $w \in R_a(v)$. (The requirement that the relations R be serial, transitive and euclidean — which is to capture belief, i.e. we say a believes ϕ at w iff for for all v with wR_av, $\phi \in v$ — ensures that beliefs are consistent, and positively and negatively introspective, respectively.)

We now want to slowly show how to adapt this notion for a probabilistic setting like ours and straight off note that some of the notions alluded

[19] Also see Bolander and Andersen [2011] for work on internal models.

[20] He does this in a plausibility (read: relational) setting so we will have to tweak his work a little bit to fit in with our probabilistic setting.

[21] Originally, this is due to Alchourrón *et al.* [1985].

to in the definition are as yet undefined for our setting. For one, we need a notion of a model M *generated by a world* $w \in M$. Second, we do not have relations defined over states but a probability function ranging over states and requiring that $R_y(w) = \{w\}$, i.e. that the root of the model be only connected to itself and only by the epistemic accessibility relation R_y of the modeller agent y. Roughly, this amounts to requiring that $\mu_y(w)(w) = 1$ in our setting. As we will later see, when taking into consideration internal *models*, in which we collect several multi agent possible worlds, this might cause problems in recapturing an actual probability distribution for an agent y that agrees with the axioms (P1)–(P3) of probability. We will, thus, speak of *pseudo-probabilistic* epistemic models whenever some of the agents have a probability distribution that does not meet the probability axioms.

Let us now turn to models generated by a world. For this, we need a notion of a *submodel*. The idea behind a submodel is that a submodel M' of a model M includes, for some subset W' of W (the set of worlds in M), all the information present at states in that subset. Let us look at the proper definition:

DEFINITION 9 (Submodel). *Let* $M = (W, \mu, t, V)$, $M' = (W', \mu', t', V')$ *be two epistemic models. We say that* M' *is a* submodel *of* M, *if*

- $W' \subseteq W$
- $V'(p) = V(p) \cap W'$
- *for all* $w \in W'$ *and* $a \in A$, $\mu'_a(w) = \mu_a(w)$
- *for all* $a \in A$, $t'_a = t_a$.

This is pretty close to the standard definition of a submodel in modal logic [cf. Blackburn *et al.* 2001, chapter 2]. We are aiming for the notion of a submodel generated by a world, that is, a model that includes all the information we can extract from that world. The third clause in the definition, then, tells us what this needs to include. Say we are considering $w \in M$, then a submodel generated by w must include not only all the worlds v where $\mu_a(w)(v)$ is defined for some agent, but also all the worlds u where for any such v $\mu_a(v)(u)$ is defined for some agent and so on. In short, we need to consider something akin to the reflexive transitive closure of the set $\{w\}$ with respect to our probability measures. Let us look at the definition:

DEFINITION 10 (Generated Submodel). *Let* $M = (W, \mu, t, V)$ *be an epistemic model. Given a set* $W_a \subseteq W$, *we recursively define*

$$W_0 := W_a$$
$$W_{n+1} := \{ \bigcup_{a \in A} \mathsf{range}(\mu_a(w)) | w \in W_n \}$$

We call the submodel $M' = (W', \mu', t', V')$, *the* submodel of M generated by W_a, *if* $W' = \bigcup_{n \in \mathbb{N}} W_n$. *If* $W_a = \{w\}$, *we call* w *the* root *of* M'.

We should recall that we only consider finite models. Now, we have already skipped the usual submodel theorem for definition 9 (i.e. to check whether a submodel really does satisfy all the same formulas for a certain subset of worlds), but we should surely check this for definition 10. This is,

however, easily seen. We already require that the generated submodel be a submodel and, thus, that any proposition $p \in \Phi$ such that $p \in w \in W'$ is also in $w \in W$. The truth-functional connectives then follow as usual. That the probabilistic information present in W_a is preserved follows simply from the construction of W' in definition 10: W', by definition, includes all the worlds in W_a, and all the worlds which any $w \in W_a$ assigns a probability to, and all the worlds those worlds assign a probability to, and so on. As the thresholds are all the same in M and M' by definition 9, also all the epistemic notions will carry over which are purely defined in terms of thresholds, probabilities and truth-functional connectives.

We now have everything in place to propose a generalization of the notion of a possible world to a multi agent possible world, *à la* Aucher [2008]. Let us look at a possible definition:

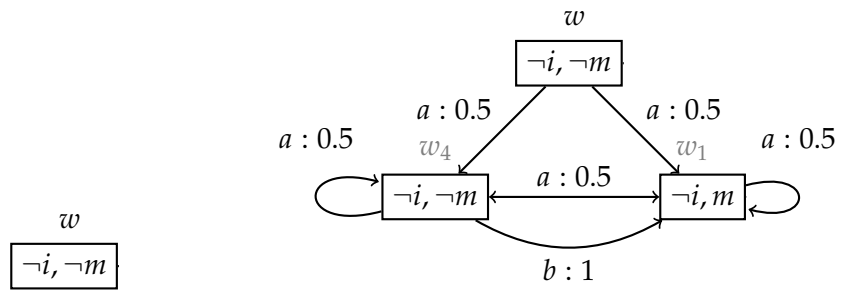
DEFINITION 11 (Multi agent possible world). *A* multi agent possible world *is a pointed (pseudo probabilistic) epistemic model* (M, w)*, where* M *is generated by* $\{w\}$ *and*

- *for all* $v \in M$ *and* $a \in A$ *where* $a \neq y$, $\mu_a(v)(w)$ *is undefined.*

This definition differs a little from the sketch we gave of Aucher's definition. As it is for Aucher [*ibid.*, pp. 18–19], a multi agent possible world is supposed to be a *generalization* of a possible world, so that in case no agent except the modeller agent y is present, the multi agent possible world becomes a normal possible world. And indeed, this is the case. Now, Aucher [*ibid.*, p. 18] also requires that the agent y have epistemic access to w, i.e. that $R_y(w) = \{w\}$. As we will see shortly, when we collect several multi agent possible worlds together to build an internal model, this is very important to Aucher. In the internal model, he requires that the roots of the multi agent possible worlds, be states that the agent y believes to be candidates to be the real world. For Aucher, however, belief is a qualitative, all or nothing notion, so that all worlds the agent considers to possibly be the real world are equally *plausible* to the agent. In our case, we want to model something far more expressive. We have degrees of belief and we want to consider every degree of belief an agent has towards a world. Certainly, taking only into account the worlds an agent may assign probability 1 to is, in the context of our project, utter nonsense.

EXAMPLE 5. Let us consider Bob this time and take it that Bob is agent y. We saw in fig. 9 that he considers a world in which Ada is sane and space travel is impossible as a good candidate to be the real world. We can see this world in fig. 11(a). Now, as we recall from fig. 9, Bob at that world also believes that Ada is indifferent as to whether space travel is possible or not. This information is not encoded in the single agent possible world in fig. 11(a), but it is present in the multi agent possible world in 11(b). Additionally, we have included that Ada, at state w_4, believes Bob is certain that w_1 is the correct state.

Let us move on and, following Aucher [2008, pp. 19–21], construct models using these new, more general worlds. Aucher [*ibid.*, pp. 20–21] defines two types of models: a type 1 model is simply a pair (M, W_a), where



(a) Single agent world (b) Multi agent world

FIGURE 11. Multi and single agent possible worlds.

W_a are the roots of the multi agent possible worlds; a type 2 model is simply a collection of multi agent possible worlds where the roots of the worlds are connected by R_y, an equivalence relation. So as to not overcomplicate things, for present purposes it will be enough to have one notion of internal models which is rather akin to the type 2 models.

DEFINITION 12 (Internal Model). *An* internal model *is a tuple* $(\mathcal{M}, W_a, \mu_y)$ *where* $\mathcal{M} = (W, \mu, t, V)$ *is generated by* $W_a \subseteq W$ *and* μ_y *is a probability distribution over* W_a.

This definition is, like definition 11, just one of numerous possible ways to go, depending also on the choices we made for 11 especially with respect to the probability assigned to the root by the modeller agent. For example, we could have also defined type 1-like models as sets of multi agent possible worlds (without an actual distribution), where the modeller agent assigns some probability to each root. For simplicity and brevity, we will only discuss the objects as defined here, as they fit best with the other definitions. We do not claim to have developed a first-person perspective here, but rather give a glimpse of what such a perspective might look like.

EXAMPLE 6. We have already seen in fig. 11(b) what the information present at w_4 in fig. 9 is like. We can now add the information present at w_1 to create an internal model for Bob, who at the real world believes w_1 and w_4 to be equally certain. We have thus, in fig. 12, modelled Bob's internal perspective as extracted from fig. 9 (again omitting 0 probability states). We can see that this comes very close to the models of definition 2, were we to delete the information regarding Ada's probabilities (and Ada's beliefs about Bob's beliefs thereby too). Were we to take 0 probability states to be expunged, we would not have the whole set of logical possibilities, but only the ones that Bob assigns a probability to. This is fine, as Bob's *internal* perspective need not include possibilities he expunges.

We can see from the example above that, as mentioned, the models considered in sections 3.1 and 3.2 can be taken as internal models as well. The addition of the real world is what makes such models objective and in the example above, the world we stipulated as represent the actual state of affairs (where Ada is insane and space travel is possible) does not even show

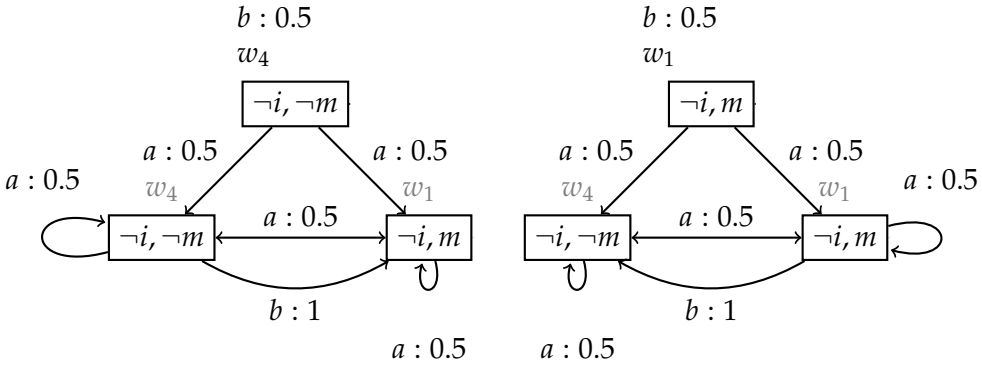


FIGURE 12. Internal model.

up (as it is assigned probability 0).[22] Thus, all notions defined simply carry over one-to-one from external to internal models, except for knowledge, of course. The question is just *where* to judge these notions. Essentially, the internal models are about the modeller agent's beliefs and his beliefs only. Certainly, other agent's beliefs play a role but only with respect to what agent y believes these agents to believe. Thus, we will only be able to talk about agent y's (possibly higher-order) beliefs in such a model and this is captured by the actual probability distribution μ_y over the roots of the multi agent possible worlds.

What the example also shows is that we can construct an internal model for an agent based on an external model. We simply take, for any pointed model (M, w_a) and agent $a \in A$, the submodel of M generated by the range of the function μ_a, $ran(\mu_a(w_a))$, at the actual world.[23]

Now, the more interesting thing for our present purposes is that Aucher [*ibid.*, pp. 27–28] manages to build an external model from a number of internal models. For this, he simply takes the set $\{(M^a, W^a)|a \in A\}$ of internal models associated with some set of agents together with a pair (w_a, V_a) of an actual world and an actual valuation. Then, he takes as the set of worlds of the external model the union of the sets of worlds of each internal model together with the actual world: for the accessibility relation he again takes the union of the internal relations and adds the pair (w_a, w^a) for each $w^a \in W^a$; the valuation, then, is simply the union of internal valuations together with the actual valuation V_a. We can do this as well:

DEFINITION 13. *Let $\{(M^a, W^a, \mu_a)|a \in A\}$ be a set of internal models over a set of agents A. The external model $\{(M^a, W^a)|a \in A\} = (W, \mu, t, V)$ is the model associated with $\{(M^a, W^a)|a \in A\}$ and a pair (w_a, V_a) is the one where*

- $W = \{w_a\} \cup \bigcup_{a \in A} W^a$,
- $\mu(a)(w) = \mu^b(a)(w)$ *for all* $a, b \in A$ *where* $w \neq w_a$ *and* μ_a *otherwise,*
- $V(p) = V_a(p) \cup \bigcup_{a \in A} V^a(p)$

(We assume the thresholds to be the same over all internal and external models.)

[22] We could have drawn it in, but that would have been less perspicuous.

[23] Aucher [*ibid.*, p. 26] does this analogously for plausibility models.

REMARK 2. This definition, we may note, opens up the possibility of a relativist perspective. Nothing in the definition tells us what the pair (w_a, V_a) ought to be, except that it is considered the real world and the real valuation of atomic propositions. It is just taken to be the *actual* way things are. We can, however, just as well consider a set of agents A, a community if you will, that agree on a set a *certainties* $\mathcal{C}$. We can then set $V(c)(w_a) = 1$ for all $c \in \mathcal{C}$, so as to build towards a real world in which all certainties of the community are (relatively) true. Most likely, this will not give us a possible world, only a partial possible world, i.e. one in which some instances of the law of excluded middle fail. Now, we can leave this as it is or decide to fill in the blanks, randomly or by admitting more and more less than certain things the agents agree on. Either way, we could thus establish a relative notion of truth, at least for some propositions. When faced with two such communities and given a picture of the *actual* actual worlds, we can simply judge the strength of each (relative) model by the intersection with reality.

Note, however, that for such a construction to be successful, there needs to be some intersection of certainties. For example, consider the example in 8, which is analogous to Wittgenstein's example of Moore and the king. An actual world relative to these two agents will turn ugly, as it will not have anything to say about Ada's sanity. Indeed, we might want to consider this relativism as a possibility only for *communities*, which we can take to be groups of agents that have the same support set.

Now, disagreement can also happen on the level of beliefs, rather than certainty. Of course, a disagreement about a certainty will imply a difference in the set of believed propositions; after all, certainties are believed with the highest degree of safety. The converse is not true. Recall fig. 10(a), where a's support set covers b's. They both share a certainty (i), but while b also believes m, a is highly confident that m is not the case.

4.3. DYNAMICS AND KNOWLEDGE-CLAIMS. As important as the epistemic notions we have seen so far are to *OC*, another important aspect there are the circumstances in which we are acting appropriately when *claiming* to know something. Utterances, in general, are incoming information. All the types of models we have seen so far are capable of modelling a particular epistemic state at a particular point in time. Incoming information will change this state. And, in fact, it is stressed in *OC* (and important to it) that these states are subject to change (recall the riverbed metaphor).

This is why a dynamic dimension is needed for a formal treatment of *OC*. In section 3 we have proposed a foundation on which we could build such a formal treatment for a single, static agent. Already sections 4.1 and 4.2 have shown that extending such a (quite expressive) formalism to a multi agent setting or outfitting it with an internal perspective leaves us with a lot of choices. In this section, the ways in which to proceed multiply again, hence we restrict ourselves to a discussion of possibly interesting approaches to dynamic probabilistic epistemic logic rather than try to define a definitive dynamics. We will go through a number of ways of approaching dynamics in our current setting but, to be clear, we will have to conclude

that none of these approaches are an intuitive fit for *OC* but may inspire a definition of such a fit down the road.

So far, we took it that an utterance of E, or incoming evidence E, results in a new probability measure at w, call it $\mu_a^{!E}(w)$ for any agent a, where

$$\mu_a^{!E}(w)(H) = \mu_a(w)(H|E).$$

This is called the *Bayesian update*. Note, however, that the Bayesian update immediately yields that $\mu_a^{!E}(E) = 1$ for the piece of incoming information E itself, unless $\mu_a(E) = 0$ or is undefined. That is to say, the new evidence, if deemed possible at all, is taken as certain after it comes in. This signals a very positive attitude towards the incoming evidence on the side of the agent. As such, it might not be the best rule to update an epistemic state with a knowledge claim, as it is not clear why an agent b should be *certain* that agent a knows p just because she said so. Of course this does not mean that the Bayesian update generates *knowledge*, what it generates is certainty.

There is a straightforward generalization of the Bayesian update due to Jeffreys [1998], which takes into account the probability we want to assign the evidence *after* it has entered:

$$\mu_a^{!E}(w)(H) = \mu_a(w)(H|E) \cdot \mu_a^{!E}(w)(E) + \mu_a(w)(H|\neg E) \cdot \mu_a^{!E}(w)(\neg E)$$

where, naturally, $\mu_a^{!E}(w)(\neg E) = 1 - \mu_a^{!E}(w)(E)$ [cf. Talbott 2015, sec. 6.2A]. This, thus, lets us set the attitude an agent will have towards a piece of evidence after he receives this evidence. It is also easy to see how taking a piece of evidence as certain in the posterior probability, Jeffreys' update collapses into the Bayesian update. However, it has "the practical problem that it requires that one be able to completely specify the direct non-inferential effects of an observation, something it is doubtful that anyone has ever done" [*ibid.*, sec. 6.2A].

Generally, the Bayesian update and Jeffrey's rule are update rules. A more elaborate update rule is found in Kooi [2003] and later in van Benthem *et al.* [2009]. Indeed, van Benthem *et al.* [*ibid.*, pp. 88–93] give a parametrized version of their rule which is capable of simulating Jeffrey's rule. None of these rules, however, seem to connect to *OC* in a meaningful way, at least in no way more meaningful than the much simpler Bayesian update.

Looking at this problem more generally, we can say that any utterance is a *model transformer*, i.e. an epistemic state model together with a piece of incoming information generates a new epistemic state. The now standard way to approach such model transformers follows Baltag *et al.* [1998], who define so-called *event models*, which are similar in structure to Kripke models but based on a set of events, some notion of epistemic access to events for agents and, instead of a valuation function, a precondition function defining what needs to be true in order for an event to occur. This approach not only has the advantage that it can model public, as well as private announcements, but also lets us take a peek inside the event structure underlying an utterance.

Van Benthem *et al.* [2009, pp. 77–8] define such an event model[24] for a multi agent probabilistic setting very close to what we have seen in subsection 4.1. Their event models consist of a set of events E, (an epistemic accessibility relation,) a set Φ of pairwise inconsistent propositions (preconditions), a function $\mathtt{pre}$ that assigns each precondition a probability distribution over E, and a function P that assigns each agent a at each event e a probability distribution $P_a(e)$ over E. In structure, this is very close to our multi agent epistemic state models. The idea here is that, in addition to the probability measures $\mu_a(w)$ of the epistemic model which give us an agents prior probabilities, what also figures into the update are: (i) "*occurrence probabilities for events* from the update model A, representing agents' views on what sort of process produces the new information" which are identified with the function $\mathtt{pre}$; and (ii) "*observation probability*, reflecting agents' uncertainty as to which event is currently being observed" which is modelled by the functions $P_a(e)$ [cf. van Benthem *et al.* 2009, p. 76]. Their update rule defines the new probability after the execution of an event model as the normalized product of the prior, observation, and occurrence probabilities.

From what we have discussed in subsection 2.3 about first- and third-person knowledge-claims, this update rule is not a very natural way to go. Of course, the prior probabilities have to play a role, as they form the "*method* of doubt and enquiry" [OC, 151] and "is the inherited background against which I distinguish between true and false" [OC, 94]. It is also clear that objective factors have to come into play when it comes to (appropriate and truthful) knowledge-claims as such a claim "must admit of being established objectively" [OC, 16]. Also our senses, our ways of observing will play a role. Whether any of these can be captured respectively by the functions $\mu, \mathtt{pre}$ and P is another question. At the very least, they have to be weighed (as can be done [cf. van Benthem *et al.* 2009, pp. 91–2]) but these weights, it seems, can change with circumstances. We can trust our senses in most cases but we are also "told in many cases that in such and such a special case you *cannot* trust them" [OC, 34]. This is not something van Benthem *et al.* [2009] account for, but this dynamic weighing would likely have to be dependent on the method, the "inherited background" for what else does an agent have to go on. To further complicate the matter, trusting ones senses is highly context-dependent. For example, in a magic show one would not trust ones sense, while when seeing a bird in a tree one can (most of the time) rely on them. Moreover, the use of seeing aids such as telescopes can play a role. This is complicated as the motivation for, as well as the level of, the trust in ones senses differ in all these situations.

Another (rather different) approach to a formal treatment of Wittgensteinian knowledge-claims is a game-theoretic approach. For example, even though sadly the details are beyond the scope of this thesis, Fiutek [2013, chapter 5] has a system in which she gives a formal version of K. Lehrer's so-called *justification game*, as presented in Lehrer [1990, p. 199 ff.] (among other works). In this game, two players, the claimant or believer and the

[24] They call it an update model, Baltag and Smets [2008a] call it action model.

sceptic engage in a dialogue over a certain bit of information (a knowledge-claim). The formal game begins with Fiutek [2013, pp. 71–3] defining so-called justification models, modelling an agent's epistemic state and her justifications which are based on an evidence structure. Justification models are based on a set of possible worlds, a valuation function and an actual world, just like a regular epistemic model. Moreover, an evidence structure[25] is defined over the set of worlds and bodies of evidence (or arguments) are ordered by inclusion. This evidence structure gives us a way to express whether a certain argument is sound or convincing and induces a relation on states. On this latter relation, epistemic notions are defined.

Getting back to the game, the sceptic gets to move first. In general, the sceptic can do one of two things: (a) he can challenge the soundness of the current argument. (b) he can challenge the current argument for being unconvincing. If he chooses strategy (a), the model is updated so as to exclude the unsound argument; if he chooses (b) the model stays the same and a new argument has to be made. Now, the claimant can reply in two ways: (a′) she can answer a challenge of type (a) with a new argument. (b′) she can answer a challenge of type (b) with a new, more convincing argument. This describes a typical round in the game. If a player cannot make any move, the player loses and the other player wins. If the critic loses, the claimant has an undefeated argument for her claim to know; if the claimant loses, her argument can be defeated and we conclude that she didn't know what she claimed to know. The technical advantage of this approach would be that the sceptic is assumed to only make a move if his move is truthful, i.e. only when the argument *really is* unsound or unconvincing. Thus, the updates here can be taken to be absolutely reliable and the only update rule needed would be the very simple Bayesian update rule.

Indeed, this seems to be a more Wittgensteinian way of handling knowledge-claims than defining some normalized posterior probability distribution. Moreover, as we have mentioned in subsection 3.3, the notion of justification defined there as a proposition implying whatever needs justification does not quite exhaust the notion of justification promoted in *OC*. The much richer evidence structure of justification models, by means of which *arguments* can serve as a justification may provide a way to include more of what counts as a justification in *OC*. In general, adding such an evidence structure to the models we have looked at so far is not problematic. What will be hard is to connect them to the epistemic operators. For Fiutek [2013], the epistemic notions, if not global, are based on evidence. In our setting, however, we would want to avoid basing the agent's probability distributions on a notion of *evidence*: "The difficulty is to realize the groundlessness of our believing" [OC, 166]. We can, if anything, thus only base the notion of knowledge on the evidence structure. The tricky thing would then be to construe the notion of knowledge in such a way that it is on the one hand based on the evidence, but on the other also influenced by

[25]The evidence structures here are akin to those of van Benthem and Pacuit [2011] and, indeed, (introspective) evidence models are just a special kind of justification model [cf. Fiutek 2013, proposition 4.2.10].

the agent's probability distribution, i.e. "the inherited background against which [they] distinguish between true and false." We can see that in any case, the merit of adding an evidence structure will not be apparently better than to leave it out.

To conclude this section, we want to point out that all of the ways we looked at above in which a model can be transformed, only the *epistemic* state is altered. That is to say, so-called *ontic facts* such as the atomic propositions that hold at a given world remain untouched. I.e. there is no way to update a model such that p is true at some w before, but not after the update. The events that these updates describe are *epistemic* events. For example, van Benthem *et al.* [2006] provide us with a simple extension that can deal with events that cause factual (ontic) change. The idea, as presented in van Benthem *et al.* [*ibid.*, section 4.1], is quite straightforward. First, they define a substitution σ to be a function from the set of propositions Φ to itself[26] such that every proposition $p_i \in \Phi$ gets assigned some proposition $\sigma(p_i) = \phi_i \in \Phi$. They call the set of all substitutions over Φ SUB_Φ. Second, they add to every event model a function $\mathtt{sub}$ that assigns each event e some substitution $\mathtt{sub}(e) \in SUB_\Phi$. Third, the model resulting from executing such an event model on a static model is as follows: the new set of worlds is the set of pairs of events and worlds (of the previous) where the world satisfies the precondition for the event. The new epistemic accessibility relation[27] connects those pairs where either two of the worlds are connected, or two of the events are connected. The genuinely new thing, then, is the valuation of this resulting model which is defined as follows: an atomic proposition p holds at a (new) state (w, e) iff w satisfies $\mathtt{sub}(e)(p)$ in the old model.[28] Thus, by, say, substituting p with $\neg p$ after some event, a new state may experience an ontic change (relative to the state it was based on). (There are other approaches to ontic change. One way similar to the above is to assign events a *postcondition*, as well as a precondition [cf. van Ditmarsch and Kooi 2008]. Another approach is to define a $flip$-operator that can flip the valuation of a formula in a model [cf. Baltag *et al.* 2008].)

[26] Van Benthem *et al.* [*ibid.*] define $\sigma : \mathcal{L} \to \mathcal{L}$ for a given language $\mathcal{L}$, but we will stick with our semantic approach from the last sections here.

[27] Note that van Benthem *et al.* [*ibid.*] work in a relational setting rather than a probabilistic one.

[28] Note that in the special case where $\mathtt{sub}$ is the identity function, which is also a possible substitution, this collapses into the standard approach of Baltag *et al.* [1998].

CHAPTER 3

Reflections

Introduction

Our first chapter was largely exegetical, our second largely formal. It is time to take a step back and look critically at what we have done. For one, we might want to examine the shortcomings the second chapter exhibits towards the first. For another we may want to have a look at further perspectives our formal approach allows for. But perhaps most importantly, we have to ask what the merit of a project such as ours is.

5. Meta Considerations

There are a number of questions we want, or need, to address at this point. (i) What is the advantage of formalizing philosophical texts? (ii) What is the point in doing epistemology, formally? (iii) Why formalize *OC*, in particular?

The inspiration of this research came largely from Achourioti and van Lambalgen [2011, p. 254], where the authors aim "to examine from the point of view of mathematical logic, Kant's formal logic and its relation to what Kant called 'transcendental logic'." What would possess them to do such a thing? What are the merits of such projects quite generally?

> To say that the Critique of pure reason is notoriously difficult is an understatement. [...] Perhaps a mathematical formalisation, however incomplete, can shed some light on [the Critique's] concepts and their relations. [...] there is hardly a better inducement to modesty than trying to come to grips with the complexities of *CPR*, not to mention the secondary literature. But the formalisation may provide a starting point.

Likewise, we tried to examine from the point of view of formal epistemology the epistemological theory put forward in *OC*. However, our aim was not necessarily to clarify. That is not to say that we have not made an effort to do so along the way. We have seen in chapter 1 the very different readings *OC* admits, e.g. differing opinions on the relation between certainty and belief or knowledge. And even though they "reflect current competing interpretations of *On Certainty*" it is not clear whether it "might be that, in spite of appearances, the responses are not that divergent" [Moyal-Sharrock and Brenner 2007, p. 3]. A formal framework, which we have by no means achieved here, can serve as a common ground for such debates.

We have shown that *OC* does admit a formal treatment — or at least aspects of it do. Moreover, the way we can treat it formally is not using

revolutionary techniques, but rather standard techniques of formal epistemology, i.e. probability theory and modal logic. How they are combined may be new, that is, do without an epistemic accessibility relation and define belief, knowledge and doubt wholly on a probabilistic structure. But new is not bad and certainly approaches like ours are gaining support due to the work of Baltag [unpublished] and Leitgeb [2014].

Our main aim in this thesis, however, was to merge two strands of research: interpretations of *OC* and formal epistemology. For one, we wanted to show experts working on *OC* that it is not a hopeless enterprise to try and formalize Wittgenstein's writing. Drawing attention to the betting interpretation of probability, I believe, we have already opened up numerous ways to formally grasp *OC*. Recall Ramsey [1931, p. 15] arguing that "the kind of measurement of belief with which probability is concerned is [...] a measurement of belief *qua* basis of action."[1] "[B]elief *qua* basis of action" does sound very much in line with *OC* and, as we have seen, the (so-called) downsides (e.g. bets on events that, if they occurred, would change the value of the pay-off) of this interpretation of probability turn out to be upsides when applied to Wittgenstein. We can argue that thus merging the betting interpretation with a Wittgensteinian view on certainties gives a more robust interpretation of probability. Moreover, it can be argued that it affirms that the thoughts presented in *OC*, when given a mathematical footing, draw a consistent picture of the situation.

This brings us to the other way of looking at this thesis. We can look at probability from a Wittgensteinian point of view, after (formally) interpreting Wittgenstein in terms of probability. Not only can we account for the odd consequence of the betting interpretation where "placing the bet may change the world, and hence your opinions" [Hájek 2012, 3.3.2] by simply pointing out that the prior probabilities make up an agent's life — "[m]y *life* consists in my being content to accept many things" [OC, 344]. But this also gives us a Wittgensteinian answer to the *problem of priors*:

> This weakness of the probability axioms generates the famous *problem of the priors*, the problem of saying where initial probabilities come from. Are they always based on evidence previously collected? If so, how does scientific inquiry get started? If instead theyre not based on previous evidence but are *a priori*, what principles govern this *a priori* reasoning? [Weisberg 2015, 1.4]

The Wittgensteinian answer would here be, I believe, that prior probabilities are neither based on evidence nor *a priori*[2] but rather a matter of convention or cultural background. The prior probability simply determines the language-game and science is but one of many games that can be played.

[1] On a biographical note, let us not forget that Ramsey, by Wittgenstein's own admission [cf. PI, preface], has influenced Wittgenstein's turn from the Tractatus to the *PI*, and thus has been an influence on Wittgenstein's later work which includes *OC*.

[2] That is, if we read '*a priori*' in the classical, Kantian way. If we read '*a priori*' simply as meaning 'prior to experience', then indeed the probabilistic priors, i.e. certainties conceived of as beliefs, are *a priori*.

> You must bear in mind that the language-game is so to say something unpredictable. I mean: it is not based on grounds. It is not reasonable (or unreasonable).
>
> It is there — like our life. [OC, 559]

It should be duly noted that we might not be content to accept the Wittgensteinian answer as convincing and launch a critique of *OC* for its failure to solve the problem of priors — or a similar problem stated in non-probabilistic terms. For example, we could argue that if the language-game is not based on *any* grounds, then it is also not based on culture, nor on community. But then how come that we *inherit* this background? Should it be possible that a child grows up, not sharing any certainties of its parents, family, or community — perhaps not even understanding their language — because the language-game is unpredictable and ungrounded?

Moreover, this (as well as the story of Moore and the king) give us a way of looking at conditioning a prior on evidence with probability 0. This usually causes technical problems and is circumvented by conventions (e.g. that $\mu(H|E) = 0$ if $\mu(E) = 0$) [cf. Baltag and Smets 2008b; van Benthem *et al.* 2009]. In our case, leaving such probabilities undefined is quite acceptable. Indeed, from an *OC* point of view, I believe, letting logic or mathematics dictate what is to happen in such cases is nonsense. A convention may, of course, be agreed on but this convention need not be the same for every agent or context (recall the context of trusting ones senses).

Quite generally, formalizing an epistemological text can lend weight to formal epistemological theories. In our case this would mean that if we have convincingly argued in section 3, the *(probabilistic) stability theory of knowledge* as constructed by Baltag [unpublished] has gained weight and it, in turn, lends weight to the stability theory of belief as found in Leitgeb [2014].

6. Is Logic Certain?

We have seen in subsection 1.2 that the dynamics of certainties, as described with the riverbed metaphor, puts the status of logic as a certainty into question. Moreover, we saw in subsection 2.4 that there is a problem concerning knowledge of tautologies: knowledge requires giving grounds, and what *ground* can I give for, say, $p \vee \neg p$? Wittgenstein [OC, 56] claims that "everything descriptive of a language-game is part of logic" but on the other hand, as we have just seen, the language-game is "something unpredictable."

> But if someone were to say "So logic too is an empirical science" he would be wrong. Yet this is right: the same proposition may get treated at one time as something to test by experience, at another as a rule of testing. [OC, 98]

Still, rules may change, right? Wittgenstein [OC, 375] recognises that "there is also something like *another* arithmetic. I believe that this admission must underlie any understanding of logic." And indeed, as any logician can attest, there a *a lot* of logics. Similarly, there are a lot of ways to construe a foundation of mathematics, e.g. constructive or inconsistent foundations. The fact that there are different logics seems to imply that an agent may

change her views on logic. Of course, such change cannot be induced by empirical means, as OC, 98 suggests.[3]

Whatever may induce such a change, this is not something our system can handle very well. We have discussed a way of expunging (classical) tautologies from the language game in subsection 3.1. This was done by defining the probability measure on a proper subset of the set of possible worlds. If we then assume that the probability of a proposition is equal to the sum of the probabilities of the states at which it holds, tautologies will not be assigned a probability. The resulting language has a probabilistic logic attached to it that is not classical. However, allowing for expunction of tautologies has consequences for the epistemic notions. For example, in subsection 3.2 we have seen that degrees of safety are normal modal operators. This is heavily dependent on the fact that probabilities are defined over the whole set of worlds. Suppose this were not the case, and assume that $H \subseteq G$ and $d_w(H) > t$, then it could be that $\mu(G)$ is undefined as we could have $v \in G$ that is outside the scope of μ. Consequently, $d_w(G)$ will be undefined and thus not greater than the threshold. Thus, we can have counterexamples to the closure of belief under implication, i.e. an agent may believe p, which logically entails q, but the agent does not believe q. This is not necessarily a bad result. In fact, it is a way to avoid logical omniscience which is considered a problem for Bayesian epistemology [cf. Weisberg 2015, 6.1].[4] However, it also has negative consequences. It would be natural to explore this system from a syntactic point of view. Any proof system we might want to build, would have to be able to do without a necessitation rule for knowledge, i.e. to infer from a *proof* of p that p is known. Any tautology would be a counter example to this rule on the semantic side.

However, intuitively proofs should generate knowledge. If we understand 'proof' as "the generation of a rational conviction of — or the rational overcoming of doubt about — the truth of its conclusion" [Wright 2004, p. 229] it should at least generate certainty (or strong conviction). *Having a proof*, however, it is also a form of justification, thus it should generate knowledge. Indeed, if one says ""I know" in mathematics, then the justification for this is proof" [OC, 563]. On the contrary, we can argue that what constitutes a proof is precisely what *OC* is concerned with (at least in the beginning), where it discusses Moore's proof of the external world.[5] And perhaps, thus, it will be quite acceptable that having a proof of p in a logical system **L** may not imply an agent's knowledge (or belief, or certainty) of p as the agent in question may, for whatever reason, not adhere to the same method of inquiry that is given by the system **L**. I.e. what constitutes a proof objectively, may not constitute a proof for the agent.

Another approach to include non-classical logics in our system we should mention is to allow for impossible and partial worlds. Priest [2006, 7.6] develops, quite straightforwardly, a probability theory on a model that is like

[3] Although this can be debated [cf. Putnam 1968].

[4] A typical technical trick to avoid logical omniscience is the use of *impossible worlds*. We will briefly discuss them further down.

[5] For a further discussion of this cf. Wright [2004].

ours, except that the valuation function can give propositions three truth values (true, false, and true *and* false). The ramifications of doing this in our current setting will have to be worked out in detail, which is sadly beyond the scope of this thesis.

7. Future Work

We have seen a number of aspects of *OC* modelled by means of formal epistemology. Most of these were mere suggestions and a lot is left open in chapter 2. For example, the right way to model utterances, i.e. the dynamics of such a system. There is still a lot of work to be done!

Clearly, a syntactic account — a *logic* of *OC* — is missing entirely, but there are many semantic issues to be worked out in order to undertake such a project. There are other things that need to be considered well before a sound and complete proof system needs to be defined.

For one, a proper dynamic machinery needs to be defined that agrees with Wittgenstein. Moreover, different types of attitudes toward a speaker uttering a claim would be helpful. In a relational setting, the following attitudes a listener can have towards an update can be defined [cf. e.g. Baltag and Smets 2008a]: an *update* $!P$ is a piece of hard information; the agent will delete all worlds satisfying $\neg P$ from her model. a *lexicographical upgrade* $\Uparrow P$ is a piece of soft information, quite reliable but not infallible; the agent will take all P worlds to be more plausible than $\neg P$ worlds, but the order within these partitions remains the same. The *conservative upgrade* $\uparrow P$ is a piece of soft information that is even less trustworthy; the agent will put the most plausible P worlds on top of all others, the rest of the order is retained. Adding such attitudes would allow us to model some other interesting situations. For example, we could take a subset M of our atomic propositions Φ that is concerned with mathematics. Then, we could model an agent a who, as a mathematician, is very trustworthy when it comes to these propositions by letting every agent have a positive attitude (say $\Uparrow$) to any utterance of a concerning propositions in M [cf. OC, 23].

Moreover, we can have agents trust their senses with respect to incoming information by assigning this kind of information a positive attitude. If we allow for the attitudes to be dynamic, we can perhaps also model an agent who comes to terms with the fact that she cannot trust her senses [cf. OC, 34]. Trust might also help model the remarks Wittgenstein makes towards how learning is supposed to work in his setting [cf. OC, 34, 263, 310–317]. There is, indeed, vibrant research going on in the intersection of learning theory and dynamic epistemic logic, see e.g. Gierasimczuk *et al.* [2014].

Conclusion

We have seen a wide variety of readings of *OC* and dismissed some of them. The foundationalist reading was deemed unsuitable because Wittgenstein so utterly diverges from traditional foundationalist views that were we to include him, foundationalism as an epistemic theory would turn out to be too broad a term. We have seen that Wittgenstein can also be interpreted in a Kantian manner, but that this is done on a metaphysical level rather than an epistemic. Moreover, Wittgenstein's and Kant's views diverge as little (if not nothing) is so absolutely stable to Wittgenstein as transcendental forms are to Kant. We have seen that Wittgenstein can be read as generating a therapeutic context in which we are to cure our (linguistic) confusions, but have argued that there is, indeed, also a theory (or a sketch of one) present in *OC*. The epistemic reading is, arguably, the one that most agrees with our reading here. Hinges, or certainties, are sure to be the most important notion in *OC*, as the title suggests. But also the semantics of knowledge-claims, disagreement about hinges and the interaction of certainty with other epistemic notions are of importance.

Taking a probabilistic approach towards formalizing these notions without them collapsing into one another owes a debt to Leitgeb [2014]. What we have identified with certainty, i.e. probability 1, has been identified in the past with both knowledge [cf. Aumann and Brandenburger 1995] and belief [cf. van Fraassen 1995]. However, as we have seen it is (in our case) too subjective to be knowledge and too strong to be belief [cf. Leitgeb 2014]. The tradition of taking belief to be probability 1 is due to the lottery paradox [cf. van Fraassen 1995], but Leitgeb [2014] has opened the door for it again (as well as for the stability theory of knowledge due to Baltag [unpublished]).

The betting interpretation of probability does, indeed, give us a good basis for modelling hinges formally. From there, we were able to define a notion of belief that gives us a subjective degree of truth and is grounded, as requested by *OC*. Moreover, we have defined a notion doubt in accordance with the conditions Wittgenstein gives us for its absence and a notion of knowledge that does not validate negative introspection. With all these epistemic notions in place, we have looked at ways of extending our system, so as to account for other important aspects of *OC*. First, we looked at a possible multi agent extension, so as to model disagreement about hinges, with the desirable result that if two agent's disagree on a certainty there is no predetermined way to persuade one another about it. Next, we took a subjective approach more in line with the presentation of *OC* and proposed one way of defining it to fit our system. Last but not least we took a look at ways to model knowledge-claims; first in terms of dynamic updates, then

in terms of a game. None of these extensions were fully developed and much work is left to do.

We have also looked at the merits of such a project and argued that this formalization sheds new light on some of Wittgenstein's remarks and some of Wittgenstein's remarks shed some light on problems existing in using formalisms like ours to model epistemic states. Moreover, we had a brief look at the possibility of expunging logical tautologies from an agent's set of certainties and ramifications of not taking logic to be certain in our system. In the last section, we mentioned some further things we could look at given our formal setting. For example, defining different attitudes towards sources of information and a connection between dynamic epistemic logic and learning theory.

In sum, we hope to have inspired a few Wittgenstein readers to do further research on formal presentations of *OC* and to have provided (or pointed towards) the tools needed to undertake such research. Likewise, we hope to have inspired the more technically-oriented reader to consider looking at probabilistic settings in a Wittgensteinian manner. Most importantly, we hope to have made both these groups curious about the other's work and opened the doors to collaborative research.

References

Wittgenstein

PRIMARY.

WITTGENSTEIN, L. [1974]. *Philosophical Grammar*. Ed. by R. Rhees. Basil Blackwell, Oxford.

— [1991]. *On Certainty*. Ed. by G. Anscombe and G. von Wright. A Blackwell Paperback. Wiley.

— [2010]. *Philosophical Investigations*. Ed. by P. Hacker and J. Schulte. Wiley.

SECONDARY.

BILETZKI, A. and A. Matar [2014]. "Ludwig Wittgenstein". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2014.

BRENNER, W. [2007]. "Wittgenstein's 'Kantian Solution'". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 7, pp. 122–141.

CANDLISH, S. and G. Wrisley [2014]. "Private Language". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2014.

CRARY, A. and R. Read, eds. [2002]. *The New Wittgenstein*. Taylor & Francis.

KOBER, M. [2007]. "'In the Beginning was the Deed': Wittgenstein on Knowledge and Religion". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 11, pp. 225–250.

MINAR, E. [2007]. "On Wittgenstein's Response to Scepticism: The Opening of *On Certainty*". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 12, pp. 253–274.

MORAWETZ, T. [2007]. "The Contexts of Knowing". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 9, pp. 165–188.

MOUNCE, H. [2007]. "Wittgenstein and Classical Realism". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 6, pp. 103–121.

MOYAL-SHARROCK, D. and W. Brenner, eds. [2007]. *Readings of Wittgenstein's* 'On Certainty'. Palgrave Macmillan.

PRITCHARD, D. [2007]. "Wittgenstein's *On Certainty* and Contemporary Anti-scepticism". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 10, pp. 189–224.

STROLL, A. [1994]. *Moore and Wittgenstein on Certainty*. Oxford University Press, USA.

STROLL, A. [2007]. "Why *On Certainty* Matters". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 2, pp. 33–46.

WILLIAMS, M. [2007]. "Why Wittgenstein Isn't a Foundationalist". In: *Readings of Wittgenstein's* On Certainty. Ed. by D. Moyal-Sharrock and W. Brenner. Palgrave Macmillan. Chap. 3, pp. 47–58.

WRIGHT, C. [2004]. "Scepticism, Certainty, Moore and Wittgenstein". In: *Wittgenstein's Lasting Significance*. Ed. by M. Kölbel and B. Weiss. Routledge.

Epsitemology

FORMAL.

ALCHOURRÓN, C., P. Gärdenfors, and D. Makinson [1985]. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". In: *Journal of Symbolic Logic* 50, pp. 510–530.

AUCHER, G. [2008]. "Perspectives on Belief and Change". PhD. Université Paul Sabatier — Toulouse III; University of Otago. URL: `https://tel.archives-ouvertes.fr/tel-00556089`.

BALTAG, A., L. Moss, and S. Solecki [1998]. "The Logic of Common Knowledge, Public Announcement, and Private Suspicions". In: *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*. Ed. by I. Gilboa. TARK98, pp. 43–56.

BALTAG, A. [unpublished]. "A Baysian Defeasibility Theory of Knowledge". Presentation Slides.

BALTAG, A. and S. Smets [2008a]. "A Qualitative Theory of Dynamic Interactive Belief Revision". In: *Logic and the Foundations of Game and Decision Theory (LOFT 7)*. Ed. by G. Bonanno, W. van der Hoek, and M. Wooldridge. Texts in Logic and Games 3. Amsterdam University Press, pp. 11–58.

— [2008b]. "Probabilistic Dynamic Belief Revision". In: *Synthese* 165.2, pp. 179–202.

— [2009]. "Talking Your Way into Agreement: Belief Merge by Persuasive Communication". In: *Proceedings of the Second Multi-Agent Logics, Languages, and Organisations Federated Workshops, Turin, Italy, September 7-10, 2009*. URL: `http://ceur-ws.org/Vol-494/famaspaper2.pdf`.

BALTAG, A., H. van Ditmarsch, and L. S. Moss [2008]. "Epistemic logic and information update". In: *Philosophy of Information, North-Holland*, pp. 361–456.

BOLANDER, T. and M. Andersen [2011]. "Epistemic Planning for Single- and Multi-Agent Systems". In: *Journal of Applied Non-Classical Logic* 21.1, pp. 9–33.

FIUTEK, V. [2013]. "Playing with Knowledge and Belief". PhD. Universiteit van Amsterdam, Institute for Logic, Language and Computation (ILLC). URL: `http://www.illc.uva.nl/Research/Publications/Dissertations/DS-2013-02.text.pdf`.

GIERASIMCZUK, N., V. F. Hendricks, and D. de Jongh [2014]. "Logic and Learning". English. In: *Johan van Benthem on Logic and Information Dynamics*. Ed. by A. Baltag and S. Smets. Vol. 5. Outstanding Contributions to Logic. Springer International Publishing, pp. 267–288.

KOOI, B. [2003]. "Probabilistic Dynamic Epistemic Logic". In: *Journal of Logic, Language and Information* 12, pp. 381–408.

LEITGEB, H. [2014]. "The Stability Theory of Belief". In: *Philosophical Review* 123.2, pp. 131–171.

TALBOTT, W. [2015]. "Bayesian Epistemology". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Summer 2015.

VAN BENTHEM, J., J. Gerbrandy, and B. Kooi [2009]. "Dynamic Update with Probabilities". In: *Studia Logica* 93, pp. 67–96.

VAN BENTHEM, J., J. van Eijck, and B. Kooi [2006]. "Logics for Communication and Change". In: *Information and Computation* 204, pp. 1620–1662.

VAN DITMARSCH, H. and B. Kooi [2008]. "Semantic results for ontic and epistemic change". In: *Logic and the Foundations of Game and Decision Theory (LOFT 7)*. Ed. by G. Bonanno, W. van der Hoek, and M. Wooldridge. Texts in Logic and Games 3. Amsterdam University Press, pp. 87–117.

VAN FRAASSEN, B. [1995]. "Fine-Grained Opinion, Probability, and the Logic of Full Belief". In: *Journal of Philosophical Logic* 24, pp. 349–377.

WEISBERG, J. [2015]. "Formal Epistemology". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Summer 2015.

PHILOSOPHICAL.

GETTIER, E. [1963]. "Is Justified True Belief Knowledge?" In: *Analysis* 23, pp. 121–123.

LEHRER, K. [1990]. *Theory of Knowledge*. Dimensions of Philosophy. Westview Press, Inc.

MOORE, G. E. [1939]. "Proof of an External World". In: *Proceedings of the British Academy*. Vol. 25, pp. 273–300.

Probability Theory

DE FINETTI, B. [1980]. "Foresight. Its Logical Laws, Its Subjective Sources". In: *Studies in Subjective Probability*. Ed. by H. Kyburg Jr. and H. E. Smokler. Originally appeared 1937 in *Annales de l'Institut Henri Poincaré*, 7, pp. 1–68. Robert E. Krieger Publishing Company.

DEMEY, L., B. Kooi, and J. Sack [2014]. "Logic and Probability". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2014.

HÁJEK, A. [2012]. "Interpretations of Probability". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2012.

JEFFREYS, H. [1998]. *The Theory of Probability*. OUP Oxford.

KEMENY, J. [1955]. "Fair Bets and Inductive Probabilities". In: *Journal of Symbolic Logic* 20, pp. 263–273.

KOLMOGOROV, A. [1933]. "Grundbegriffe der Wahrscheinlichkeitrechnung". In: *Ergebnisse Der Mathematik*.

RAMSEY, F. [1931]. "Truth and Probability". In: *The Foundations of Mathematics and Other Logical Essays*. Ed. by R. Braithwaite. Originally appeared 1924, electronic edition from 1999. Page numbers cited are from the 1931 publication. Kegan et al. Chap. 7, pp. 156–198.

VINEBERG, S. [2011]. "Dutch Book Arguments". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Summer 2011.

Others

ACHOURIOTI, T. and M. van Lambalgen [2011]. "A Formalization of Kant's Transcendental Logic". In: *The Review of Symbolic Logic* 4.2, pp. 254–289.

AUMANN, R. and A. Brandenburger [1995]. "Epistemic Conditions for Nash Equilibrium". In: *Econometrica* 63.5, pp. 1161–1180.

BLACKBURN, P., M. de Rijke, and Y. Venema [2001]. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

KANT, I. [1998]. *Critique of Pure Reason*. Ed. and trans. by P. Guyer and A. Wood. Cambridge University Press, UK.

KRAUSZ, M. [1984]. "Relativism and foundationalism: some distinctions and strategies". In: *The Monist* 67.3, pp. 395–404.

MCGRATH, M. [2014]. "Propositions". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2014.

PRIEST, G. [2006]. *In Contradiction. A Study of the Transconsistent*. 2nd ed. Oxford University Press.

PUTNAM, H. [1968]. "Is Logic Empirical?" In: *Boston Studies in the Philosophy of Science* 5.

ROWE, W. and N. Trakakis [2007]. *William L. Rowe on Philosophy of Religion: Selected Writings*. Ashgate Contemporary Thinkers on Religion: Collected Works. Ashgate.