DYNAMIC INTROSPECTION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Michael Cohen

August 2021

Abstract

In this dissertation, I study an under-appreciated type of introspection that helps explain how we can come to know the world around us. Much of today's epistemology, the philosophical study of knowledge, can be traced back to the work of René Descartes, and to the skeptical challenge he posed: what can we know beyond doubt? A longstanding response to Descartes' challenge is to locate the underlying assumptions that drove him, and others, to their skeptical conclusion; Descartes reached his skeptical position while assuming that subjects have clear and direct access to their own minds, that they have full introspection. During the last 70 years, philosophers have challenged these assumptions, developing tools in epistemology and philosophical logic to study the kind of ignorance subjects have about their own knowledge.

Nevertheless, there is a type of introspection, which I call *dynamic introspection*, that has not caught the attention of epistemologists. This introspection is about, on the one hand, what we know about the way new experience will affect our future knowledge (*where am I going from here?*) and on the other hand, what we know about the ways we got our current knowledge (*how did I get here?*). Many epistemological debates either ignore such questions or implicitly assume that we are always certain about these matters. Do I need to have the prior certainty that my sources of information are trustworthy in order to gain knowledge from them? The answer to such questions depends on the dynamic introspective abilities we assume we have. Understanding the nature of dynamic introspection can vindicate our reliance on ordinary knowledge in the face of skeptical doubt. A subject might know that there is a tree in front of her (as a result of her perceptual experience) without having the ability to explain the exact source of her knowledge. I argue that this lack of dynamic introspection is natural, and does not lead to skeptical problems.

In this dissertation, I introduce a new logical framework for analyzing and modeling dynamic introspection. After introducing the logical framework in Chapter 2, I use it to analyze important debates from the contemporary epistemological literature in which, I argue, dynamic introspection principles implicitly play a key role, and much can be learned by making them explicit. One debate concerns the skeptical challenge to the possibility of our perceptual knowledge about the external world: how can we answer the skeptic who doubts the possibility of perceptual knowledge? The second debate concerns the limits to which we can know and access our own internal mental world:

what can we know about our own knowledge? By making the underlying dynamic introspection principles explicit in such debates, I hope to make a meaningful contribution to our understanding of our own knowledge.

Studying dynamic introspection has two interrelated dimensions: logical and epistemological. I believe that although the importance of dynamic introspection principles in epistemology can be appreciated without an understanding of their underlying role in epistemic logic, a greater understanding is achieved by grasping their logical manifestation.

Chapter 1 of this dissertation introduces and surveys the epistemological and formal themes that play a role throughout the dissertation, including the epistemological problems of perceptual skepticism and bootstrapping, the distinction between externalism and internalism in epistemology, the philosophical uses of epistemic logic, and the role of dynamic introspection in epistemology.

Chapter 2, *Opaque updates*, is devoted to a logical analysis of dynamic introspection principles. The chapter outlines a logical system that can represent epistemic learning events (updates) that are not transparent to the epistemic agent, but opaque. In such updates, the agent lacks dynamic introspection. The chapter highlights the importance of two dynamic introspection principles in particular: the no-miracles and perfect-recall principles. These principles are taken as axioms in many systems of dynamic epistemic logic, and they break apart when updates are not transparent. Studying the epistemological role of these two principles guides the next two chapters of the dissertation. A version of Chapter 2 was published as an article in *Journal of Philosophical Logic* in 2021, under the title *Opaque updates*.

Chapter 3, *The problem of perception and the no-miracles principle* offers a novel analysis of a standard skeptical argument about knowledge gained by perception. I argue that the no-miracles principle is implicitly assumed in the skeptical argument. The skeptic can be understood as making a very strong introspective assumptions, viz. accepting the no-miracles principle, at the price of losing perceptual knowledge. The chapter describes a different philosophical picture, one in which dynamic introspection is lost, while perceptual knowledge is not threatened. A version of Chapter 3 was published as an article in *Synthese* in 2020, under the title *The problem of perception and the no-miracles principle.*

In Chapter 4, *The bootstrapping problem and perfect recall*, I study the importance of the perfect recall principle in the context of the bootstrapping problem. While the perceptual skeptic purports to show *a-priori* that perceptual knowledge is impossible, the bootstrapping problem seems to show that unacceptable philosophical conclusions follow *posterior* to a learning event, at least for some philosophical theories. I argue that bootstrapping reasoning implicitly assumes the dynamic perfect recall principle, and that such reasoning can be blocked by abandoning the latter principle. Together, chapters 3 and 4 paint an epistemological picture in which both no-miracles and perfect-recall fail, but ordinary knowledge is free from skeptical paradoxes.

In Chapter 5, *Inexact knowledge and dynamic introspection*, I offer an analysis of inexact knowledge, the knowledge we get from our imperfect perceptual faculties. I argue that inexact knowledge is the result of inexact observations, which is a type of opaque update. My analysis of inexact knowledge challenges an influential argument in contemporary epistemology, according to which inexact knowledge is incompatible with the KK principle (knowledge entails knowledge of knowledge). I argue that inexact knowledge is incompatible with dynamic introspection, not with the KK principle. A version of Chapter 5 was published as an article in *Synthese* in 2021, under the title *Inexact knowledge and dynamic introspection*.

Although the contents of the chapters are very much connected, I wrote each chapter so it can be read without prior knowledge of the others. The logical framework presented in Chapter 2 is very useful for understanding dynamic introspection, but it is by no means necessary.

Acknowledgments

The ideas for this dissertation were developed during my regular meetings with Krista Lawlor and Ray Briggs in my third year in the PhD program. In these meetings, the main argument of Chapter 5 slowly took form. Once that argument was established, it gradually became apparent that a larger project can be developed from it, going beyond the particular details of Chapter 5. I am immensely grateful to Ray and Krista for the countless hours spent in these meetings, and for enabling the perfect environment for developing this PhD project. Of course, I could not imagine coming up with these ideas—connecting dynamic epistemic logic and epistemology—without Johan van-Benthem's consistent mentoring, encouragement and insights during my entire time in the program. Krista, Ray, and Johan significantly helped me develop, write, and revise the dissertation (especially Chapters 2, 3 and 5 that turned into journal papers), teaching me how to take a philosophical idea and transform it into a publishable paper, how to respond to reviewers, and how to present these ideas in talks. I could not have asked for a better and more supportive dissertation committee. Beyond the proper dissertation work, Krista, my advisor from day 1 at Stanford, was always available for questions, comments, suggestions, advice, tips, and mentoring, helping me in any aspect of PhD student life at Stanford. Thank you, Krista for your mentoring. Ray and Johan also deserve special thanks for building a sense of community among the students in the department: Ray's weekly group meetings, and Johan's Spring seminars and the CSLI workshop organisation (and the Spring hikes!). This student community was pivotal for developing and presenting my work in progress throughout the years: thank you for facilitating such a community.

I would like to thank the friends who helped me with my dissertation project throughout the years in the department: Adam, Shane, Peter, JT, Nick, Nathan, Adwait, John, Johnathan, Wang, Grace, Rob, Chris, Michael, Steven, Francesca, Declan, Dave, and Hanna. Our discussions, in GSW, over drinks, or while sharing office spaces, greatly helped my dissertation work.

Parts of this dissertation were presented in the following workshops and conferences: the 2018 NASSLLI student session (Pittsburgh, June 2018), the 2019 Pacific APA (Vancouver, Canada), the Masterclass in Theoretical Philosophy with Timothy Williamson (Tubingen, July 2019), the Formal Epistemology Workshop (Turin, 2019), the Glasgow Graduate Conference in Epistemology and Mind (May 2019), and the Knowledge and its limits at 20 conference (Geneva 2020). I would like to thank

all the participants and commentators in these events for their helpful comments. I would also like to thank the anonymous reviewers who commented on the paper versions of the chapters of the dissertation.

I would like to thank Thomas Icard, for the many meetings I had with him, and his willingness to participate in the oral defense, and for Daniel Lassiter for agreeing to chair the oral defense.

Finally, I would like to thank my family for supporting me from afar in my academic journey, and for my wife Noe, who made this entire journey genuinely meaningful: Noe, I could not have done this without your constant support. Thank you for sharing this journey with me, and for your encouragement in every aspect along the way.

Contents

List of Figures

Chapter 1

Introduction

Chapter abstract: In this chapter, I offer an overview of a number of epistemological themes that play an important role throughout the dissertation. The chapter aims to provide the conceptual background to the work presented in later chapters of the dissertation, and to situate my epistemological starting position and perspective. I start by motivating an old epistemological challenge: the problem of getting to know the external world via the sources of information available to us, such as our perceptual faculties. The distinction between internalist and externalist theories of justification, knowledge, and evidence is sketched, with a focus on developments within externalism, including naturalistic epistemology and knowledge first epistemology. A philosophical introduction to epistemic logic and its connections to epistemology is then provided. Epistemic logic allows for the study of introspection principles in a precise manner, and have led to renowned investigation between externalism and introspection. I contrast the existing literature on introspection in epistemic logic, which I call static introspection, with an understudied form of introspection, dynamic introspection.

1.1 Knowledge of the external world

Understanding our knowledge of the external world is a key task of epistemology. As a rough starting point, consider the following, very simple, model of knowledge of the external world: there is some epistemic agent, and some source of information. The epistemic agent can be anything that is able to store information about the external world; this presumably involves the ability to represent the external world. Humans are epistemic agents. The source of information can be anything that can provide information about the external world; perceptual faculties are a prime example in epistemology. The source of information provides the epistemic agent with information about the external environment and, if all goes well, the agent receives the information and comes to know it. For example: when I look at my phone's weather app to check that today will be sunny, I am the epistemic agent, the weather app is the source of information, and the information about the external world is that today will be sunny. After checking the app, which provides me with the information that today will be sunny, I come to know that fact.

In this dissertation, I focus on the *dynamic* aspect of this simple model. The model describes a dynamic process, with three distinct elements: the initial stage, before the agent receives information, the event of the source providing the information (the learning event), and the resulting stage of coming to know the information. While this process seems very straightforward, tracking what the agent knows *at and about* each different stage of this process can be quite intricate, as this dissertation aims to show.

Sometimes, things do not go well. The information in the weather app has not been updated with the latest weather data, and provides the wrong forecast; our perceptual capacities are prone to biases, mistakes, and misperceptions. In general, our sources of information about the external world are fallible. To avoid this hurdle, we would like to make sure that our sources of information are reliable. To do so, we can further investigate the mechanisms of our sources of information (has the app's data been updated?) or compare it to different, independent, sources of information (what do other weather apps say?). But sometimes, it seems, it is impossible to verify our sources of information (at a given moment, or in general), and many debates in epistemology center on such scenarios.

A prime example of such a scenario is known as *the problem of perception.* If, as some philosophical traditions have argued, all of our knowledge of the external world comes from our perceptual faculties, how can we verify that our perceptual faculties do not mislead us? The problem here is that in order to verify that our perceptual faculties are reliable, we cannot rely on those very same faculties. If perception is our only source of information, then in order to verify that our perception is accurate, we will need to use perception. But using our perceptual faculties to justify our perceptual faculties seems circular. The problem is to resolve, or explain away, this apparent circularity. The problem of perception is presented for our perceptual faculties, but a similar problem applies to any source of information such that the only way to justify the reliability of the source is by using that very same source (memory is another one: how can we rule out that the world was created, with all of our memories, just 5 seconds ago?).

According to the skeptical response to the problem, we don't really have the perceptual knowledge we think we have. Versions of this core argument have played a critical role in reshaping modern epistemology, going back to Rene Descartes' *Meditations* (Descartes, 1985) and to Michel de Montaigne's *Essays* (de Montaigne, 2007). A full articulation of the problem of perception (and responses to it) requires much more attention to the philosophical assumptions and argument structure at play. I offer such articulation in Chapter 3 of this dissertation (*The problem of perception and the no-miracles principle*). Here I want to point out how the problem can be viewed from a dynamic perspective.

Recall the distinction between the initial stage, the learning event, and the resulting stage. Suppose that in the initial stage, the epistemic agent does not know if their perceptual faculties (considered as a source of information) are reliable. The agent therefore does not know if the information they will receive from the source is truthful and therefore results in knowledge of the external world. So, at the initial stage (even before receiving the information) the agent might have reasons to believe that as a result of the learning event, they will not gain knowledge of the external world. The question is whether ignorance about the reliability of the information source (perception) at the initial stage makes it impossible for the agent to gain knowledge at the resulting stage, and if so why. Answering this question one way or the other involves the commitment to dynamic principles that connect the different stages of the learning process, and the agent's knowledge about those stages. I call such principles dynamic introspection principles, and I articulate their philosophical importance in the body of the dissertation.

There are many different ways to approach the problem of perception. In the next section, I sketch the distinction between externalist and internalist theories in epistemology, and the ways they approach the problem of explaining our knowledge of the external world. This will help clarify the methodological framework of this dissertation.

1.2 Externalism and Internalism in epistemology

Within philosophy, the terms externalism and internalism are used for a variety of distinct, but related debates in different fields of philosophy, including in philosophy of language and mind, philosophy of action, and epistemology. Although the exact meaning of the distinction varies, at a rough and general level, externalist and internalist disagree, for a given phenomenon, whether that phenomenon solely supervenes on the mental states of a given agent, or whether it further supervenes on the environment external to the agent. Normally, the phenomenon in question is considered, traditionally or intuitively, to be internal to the agent, and externalist arguments challenge this commitment.

For instance, in philosophy of language, externalists about meaning argue that meaning is partially determined by external factors, not just by the mental state of the agent ("Meaning just ain't in the head" Putnam, 1975: 227); internalists disagree. Externalists about mental content further argue that the *content* of mental states like beliefs is determined by external factors (Burge, 1979). Externalism about moral motivation holds that moral judgments (an internal mental state) are not necessary for moral motivation (Rosati, 2016). These types of debates naturally raise the question of the boundary of one's mind: debates about the extended mind and embodied cognition (Clark and Chalmers 1998) can shift the line between what is internal to the agent, and what is not.

In philosophical methodology more broadly, externalism favours the third person perspective as the starting point of philosophical investigation; internalism favours a first person perspective. The different perspectives affect the role of intuitions in philosophical investigation, for instance. Due to this methodological focus, positions that stress the third person perspective are sometimes called meta-externalism (e.g., Cohnitz and Haukioja 2013).

In epistemology, the distinction between externalism and internalism often concern the epistemic concepts of justification, evidence and knowledge. According to internalists about justification, whether or not an agent is justified in their beliefs is solely determined by their mental states: if two agents differ in their justification towards some belief, then there must be a difference in the mental states of the two agents. Internalism about justification is further divided into various theories, of which I mention *access internalism* and *mentalism*. According to access internalism, justification is directly accessible to the epistemic agent (the agent can be always aware of their justification). Mentalism is weaker than access internalism, and only requires that justification is wholly determined by the agent's internal mental states (whether or not the agent has access to them) (Feldman and Conee 2001). For access internalism in particular, the connection between justification and the introspective abilities of the agent is very important. Having justification is intimately connected to the ability to reflect on one's own mental states. Although access internalism is less popular in contemporary epistemology, it has been historically quite influential. The Cartesian requirement that ideas must be clear and distinct in order to count as secure from doubt (Descartes 1985) assumes that the agent has the ability to reflect on their own ideas, and that such reflection is crucial for attaining knowledge.

Externalist views on justification deny the internalist assumption: such views hold that factors external to the agent's internal mental states can determine whether an agent is justified in their beliefs. According to externalists, it is possible for two agents to have the exact same internal mental states, such that one agent has justification for their beliefs, while the other doesn't, due to external factors in the two agents' environments. One prominent example of externalism about justification is *reliabilism*. According to reliabilism, an agent is justified in their beliefs if the processes that generated the beliefs are epistemically reliable (Goldman 1979). Causal theories of justification, which require that a justified belief was formed by the right causal processes are another example of an externalist theory. There are many varieties of externalism about justification, but for the purposes of this dissertation, these differences can be mostly ignored. We understand externalism in a negative manner: externalism does *not* assume that the mental states of the agent wholly determine the epistemic ones.

Although externalism plays an important role in each chapter of this dissertation, the notion of justification does not. The extended analysis of justification in 20th century analytic epistemology largely stems from the (now mostly abandoned) project of reductively analyzing the concept of knowledge into elements that include some notion of justification. This historical fact explains the extensive focus on justification (and adjunct concepts such as *warrant*) in debates about externalism and internalism in epistemology. But these debates have extended beyond justification.

Externalists about evidence hold that whether or not an agent has evidence for some proposition is partially determined by the external environment; internalists take evidence to be an internal mental state. According to some externalists, evidence is factive: for something to count as real evidence it has to be a true piece of information (Williamson (2000), Littlejohn (2013), Bird (2018)). If evidence is determined by external factors, it becomes very clear why two mentally identical agents might have different justified beliefs: they just don't have the same evidence (even if they think they do). Variants of externalism and internalism about justification naturally transfer to debates about evidence, although debates about evidence are not as tightly connected to the analysis of knowledge project.

Finally, there is externalism and internalism about knowledge. Since, within analytic philosophy, there is a large consensus that knowledge is factive, the position of externalism about knowledge might seem confused. If knowledge is factive, then, necessarily, the state of knowledge is sensitive to the external environment that goes beyond the agent's internal mental state. Clearly, knowledge involves external factors in that sense, and there isn't much debate about that. Externalism about knowledge is rather the position that factors that distinguish knowledge from mere true belief might be external, and so inaccessible, to the agent. Likewise, internalism about knowledge holds that any difference between mere true belief and knowledge involves an internal mental state for the knowing agent.

Since epistemological externalism does not require the agent to have any kind of access or mental connections to the external factors that guarantee knowledge (or justification), externalist positions are often interpreted as limiting the introspective abilities of epistemic agents (as less internal access to the underlying factors implies less ability to reflect on those factors). However, making explicit this lack of introspection can take different forms. In this dissertation, I offer a novel way to think of epistemic introspection, from a dynamic perspective. This leads to new ways of understanding externalism about knowledge.

The division between externalism and internalism in epistemology offers two different ways of responding to the problem of perception, and the existence of our knowledge of the external world more generally. The two ways do not necessarily correlate with externalism and internalism, nor do they exhaust the ways of responding to the problem. Nevertheless, it is instructive to sketch some natural responses to the problem according to the externalism internalism divide.

Recall that according to the problem of perception, it seems that we are not always in a position to know that our perceptual sources of information are reliable. The skeptic argues that the ignorance we have about our sources of information leads to ignorance about the external world: in order to really know the external world, we have to be justified in believing that our sources of information are reliable. But we don't have such justification. Therefore, argues the skeptic, we don't really have knowledge of the external world.

Here are two (non exhaustive) responses to this skeptical conclusion: the first is to deny that we lack any kind of justification that our perceptual sources of information are reliable. The second is to deny that such kind of justification is actually needed in order to gain knowledge of the external world. Externalist views (about knowledge, evidence and justification) tend to take the latter response. Externalists about perceptual knowledge hold that in order to gain perceptual knowledge, certain external factors must hold for the source of information (e.g. right causal connection to the object of perception). Externalists are *not* further committed to the claim that the agent must be aware of these factors; the agent might be completely ignorant about them. Since, according to externalism, the agent is not required to have this further level of justification that their sources of information is in fact reliable, the immediate externalist response to the skeptic is to reject the assumption that not knowing that one's source of information is reliable implies not having knowledge.

Internalists about perceptual knowledge (and especially access internalists) tend not to accept such reply to the skeptic. Instead, they might argue that the agent does have some justification for the reliability of their sources of information. According to one internalist position, perceptual justification comes, *prima facie*, with the perceptual experience (Pryor 2000). According to another broadly internalist route, we have some sort of epistemic entitlement to rule out the scenario raised by the skeptic (Wittgenstein 1969, Wright 2004). The nature of this type of entitlement is often quite elusive, and will not be at the focus of my work. Nevertheless, I consider such views internalist as they do not rule out the possibility that agents are able to access and reason about this type of entitlement, from their internal perspective.

It is instructive to classify these two different responses to the skeptic as *modest*, or *weak foundationalism* (Lyons, 2016). Foundationalism is one response to the regress problem; the regress problem is the problem of making explicit the justificatory structure of our knowledge and belief. Assuming that any piece of knowledge requires further justification seems to lead to infinite regress in the justification chain. According to foundationalism, the chain (or tree) of justification has a root: pieces of knowledge or belief that are considered basic. According to modest foundationalism some beliefs (or knowledge) just do not require other beliefs for support (Lyons, 2016). The different responses to the problem of perception and their connection to modest foundationalism are explored in more detail in Chapter 3.

This dissertation (especially chapters 2, 3 and 4) develops a novel externalist epistemological position, which allows the combination of the agent's ignorance about their sources of information with the agent's knowledge of the external world. In simple terms, the agent can know without knowing *how* they know: agents might be ignorant about their sources of information, and the ways they came to know what they know. In particular, they can have ignorance about the dynamic processes that generate their knowledge. This is externalism in *form*, not in *content*: I do not advocate for a particular external *thing* (e.g. reliable process, causal structure) needed for an externalist conception of knowledge, rather I articulate the formal structure and commitments of any broadly externalist theory. I present a way to conceptualize an agent who does not have the type of dynamic introspection needed to answer questions like *how do I know the things I know*. I show how to logically represent and reason about such ignorance in Chapter 2, and the philosophical implications of such dynamic ignorance in Chapters 3 and 4. I argue that epistemic agents lack a dynamic form of introspection, and that this form of introspection has been ignored in the epistemological literature.

The next two subsections survey two prominent positions within externalism, that play some role throughout the dissertation.

1.2.1 Naturalistic epistemology

Naturalistic epistemology denotes a family of views that consider epistemology as continuous with natural sciences, and that aligns the methodology of epistemology with a broader naturalistic philosophical methodology (Quine 1969). Naturalism is not necessarily the most dominant methodology in epistemology (this is in contrast to contemporary philosophy of mind, for instance).

Naturalistic epistemology is not necessarily wholly externalist. One can hold that epistemology, as a discipline, essentially reduces to cognitive psychology, and then argue (using methods from cognitive psychology) that epistemological concepts like belief and justification are accessible to epistemic agents. This position combines naturalistic epistemology in methodology with access internalism about belief and justification.

One clear manifestation of externalist naturalistic epistemology is the view that knowledge, like gold, is a natural kind (Kornblith 2002). According to this view, knowledge is something that both humans and non-human animals have, and it plays an important role in causal explanation of the natural world. Knowledge, in this sense, is directly connected to evolutionary processes. Attributions of knowledge to an organism about its environment essentially describe a certain fit between the organism and its environment, a fit that is to be expected due to natural selection. Saying that Salmon know to distinguish their predators from their prey attributes knowledge to fish, one that is to be expected from an organism for its survival. If salmons did not have a reliable epistemic method to make this distinction, how would they survive? Epistemic abilities are the natural result of evolutionary processes (see Bradie and William 2020).

The connection between this kind of naturalistic epistemology and externalism is very tight. I know my way to the closest supermarket. Migratory birds know their way to their breeding grounds. In this case two widely different organisms share the same kind of state (the knowledge state). Since the cognitive and neural wetware of the two organisms is so radically different, the common kind of knowledge state they share must be articulated, at least partly, by an appeal to functional or behaviouristic elements (both I and the migratory birds consistently reach our respective goals). Such elements are observable and therefore in an important sense external. An internalist epistemology, both of the access type and of the weaker mental supervenience type, would find it much harder to find the commonality between myself and the birds (asking the birds to justify their ways will not work).

As the talk of functional nature of mental states suggests, I view naturalistic epistemology as closely connected to the on going attempt in philosophy of mind to offer a compelling naturalistic picture of the mental. Naturalistic epistemology is criticised as lacking the ability to explain the normative nature of epistemology (Kim 1988). The criticism is serious, but answering it involves a much larger project of reconciling naturalism with normative thought in philosophy (and in meta-ethics in particular). The problem of finding a place for normativity in a naturalistic world view is not special to epistemology.

Although this is not always made explicit, the epistemological views presented in this dissertation favour a naturalistic view of epistemology.

1.2.2 Knowledge first epistemology

Another prominent externalist view within epistemology is *knowledge-first epistemology*, introduced and defended by Williamson (2000). Knowledge-first epistemology holds that the notion of knowledge has a priority over the notion of belief. This priority, according to knowledge-first epistemology, is manifested in many different ways.

The most obvious one is conceptual priority. The project of analyzing the concept of knowledge in terms of a true belief with some special type of epistemic property was one of the central projects of analytic epistemology in the latter half of the 20th century. Williamson (2000) argues that that project assumed the priority of belief (belief-first epistemology) without much argument. Conceptually, according to Williamson, it makes more sense to take the success concept (here knowledge) as primitive, and understand its less successful relatives (like mere belief) in its terms. It is much easier to understand an imperfect circle in terms of an almost perfect circle, than to understand a perfect circle by ruling out all the ways a circle can be imperfect (2000: 4). Likewise, we take memories to be factive in everyday use; we are also aware of the concept of false memories. We understand the concept of false memories in terms of its being almost like our regular conception of memories; we don't think of normal memories as a special type of false, or non-factive type of memory: memories have a conceptual priority over false memories.

Knowledge-first epistemology *uses* the concept of knowledge to analyze other epistemic concepts and norms, including the concepts of belief, justification, and evidence, and norms related to assertion and action. To name just a few uses of the concept of knowledge in knowledge-first epistemology, believing p is understood as treating p as if one knows p (2000:47); evidence is taken to be *identical* to knowledge; the states of having justification for p or having p as part as one's evidence requires knowing p; a necessary condition for asserting p or acting on the basis of p is that one knows p. Since so many central epistemic notions are understood in terms of knowledge, knowledge-first epistemology is bound to be thoroughly externalist: the right belief, justification, evidence, and action require knowledge, and knowledge requires truth, i.e., a match with the external reality.

Knowledge first epistemologists appeal to other aspects of the priority of knowledge over belief that go beyond the conceptual, a-priori, realm. Knowledge is more central in natural language than belief: the verb *to know* is, cross-linguistically, statistically much more common than corresponding non-factive verbs such as *to think* or *to believe* (Hansen et al. 2021); the concept of knowledge also has *developmental priority* over belief: young children master the use of the concept of knowledge before they master the concept of belief (Nagel 2013). These empirical finding suggests that the concept of knowledge is more 'user-friendly' than that of belief. This ease of use is also reflected in formal representations of knowledge and belief: knowledge is a simpler concept to characterize formally (as a factive, non-gradable mental attitude). As section 1.4 shows, within epistemic logic, it is quite common to *explicitly* define belief in terms of knowledge (as the epistemic possibility of knowledge); defining knowledge in terms of belief is not common at all in epistemic logic (cf. Halpern et al. 2013, Bjorndahl 2020).

Methodologically, and following knowledge first epistemology, this dissertation prioritizes knowledge over belief. Knowledge is taken as a primitive notion, which is used in epistemological investigations, both conceptually and formally (via the use of epistemic logic). The focus of this dissertation is on the 'knowledge' formulation of epistemological issues: perceptual *knowledge*, introspection that involves *knowledge*, externalism about *knowledge*, and the bootstrapping problem applied to *knowledge* (as opposed to perceptual *beliefs*, introspection that involves *beliefs*, etc.). The notion of belief plays a marginal secondary role in the dissertation, and it is not assumed that knowledge should be analyzed as a type of belief. These methodological commitments should *not*, however, be conflated with all of the epistemological views presented in Williamson (2000).

1.3 The Bootstrapping challenge

A major objection to epistemological externalism, and more generally to modest foundationalism, is known as the *bootstrapping* or *easy knowledge* problem. The problem was first formulated by Fumerton (1995) and Vogel (2000, 2008) as an objection to reliabilist theories of knowledge and justification. Cohen (2002) formulated a similar problem that extends to modest foundationalist positions that involve *basic knowledge*, i.e. the possibility of getting knowledge from a source of information without knowing that the source is reliable. White (2006) applies the bootstrapping reasoning against *Dogmatism* (Pryor 2000, 2004), the internalist position that one can gain prima-facie justification from perceptual appearances without being antecedently justified that these appearances are veridical. Weisberg (2010), Titelbaum (2010) and van Cleve (2003) argue that the problem is more general and affects a wide range of epistemological views.

In the face of the problem of perception, recall that the skeptic argues that the epistemic agent does not have perceptual knowledge, since the agent does not know that their sources of information are reliable. The externalist agrees that the agent might not know that their sources of information are reliable, but argues that this does not imply the agent does not have perceptual knowledge. In other words, the externalist tries to draw a modest epistemological picture, in which the agent has some ignorance about the environment (ignorance about the reliability of the source of information), which does not imply skepticism. The externalists point is to offer a middle ground which rejects skepticism on the one hand, but does not suppose that the agent has full knowledge of their environment.

According to the bootstrapping challenge, this modest, middle ground position cannot be maintained. The externalist position seems to further allow the agent to learn that their sources of information are reliable, without the proper evidence to come to this conclusion. This illicit form of learning, bootstrapping, seems to follow from externalist theories. The objector concludes that externalist theories should be rejected on that ground.

To see how bootstrapping can occur, consider the following example. Suppose that the agent is looking at a clock in order to come to know the time. Assume that the agent does not initially know that the clock mechanism is reliable. According to the skeptic, since the agent does not know that their source of information is reliable, the agent cannot come to know the time as a result of looking at the clock. Moreover, the skeptic assumes that the agent is able to reach this conclusion even before the learning event (looking at the clock) occurs. According to the externalist, the fact that the agent does not know that the clock is reliable at the initial stage does not imply that the agent does not come to know the time by looking at the clock. At the initial stage, the agent did not know the time and did not know that the clock is reliable. At the resulting stage, the agent knows the time. Does the agent know that the clock mechanism is reliable as a result of looking at the clock? Intuitively, the answer should be no, since the agent did not gain any evidence about the reliability of the clock just by looking at it. However, according to the bootstrapping objection, the externalist framework allows the agent to bootstrap their way to the conclusion that the clock must be reliable.

How exactly does externalism allow for bootstrapping reasoning? This process is not always made entirely explicit in the literature, and often assumed to invoke inductive reasoning (Vogel 2000). The problem is that this tries to explain one mysterious reasoning process (bootstrapping) with another form of reasoning plagued with epistemological riddles (inductive reasoning). It is also not clear in the literature if induction is essential for bootstrapping reasoning (Titlebaum 2010). In Chapter 4, I use the dynamic framework I develop in this dissertation to make the bootstrapping reasoning explicit as possible. I argue that bootstrapping reasoning implicitly assumes a substantial dynamic introspection principle that externalist theories have independent reasons to reject. In short, since the externalist epistemic agent does not always know how they know (at the resulting stage), there is no reason for the externalist to accept the substantial dynamic introspection principle that the bootstrapping reasoning requires. If this is correct, then the bootstrapping challenge does not threaten externalism in epistemology.

1.4 Epistemic logic and its connections to epistemology

In the dissertation, epistemic logic is used to formulate and investigate the core tenets of externalist epistemology, and the precise structure of the bootstrapping and skeptical problems. In this section, I go over the basic structure of epistemic logic, with an eye to clarifying its use in epistemology, and addressing some common misconceptions about it.

Although epistemic logic has its roots in medieval philosophy (Boh 1993), its modern inception tracks back to Jaakko Hintikka's work *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (1962), which was itself a part of a larger prolific period of research in modal logic during the mid 20th century (Ballarin 2021). Hintikka's work has established or influenced many of the uses of—and debates about—epistemic logic in epistemology, including: the principled relation between knowledge and belief, the closure of knowledge under logical entailment, the nature of introspection principles in epistemology, the analysis of knowledge-wh (knowledge followed by a wh-question) vs. knowledge-that, and *Moore's paradox* regarding the nature of self contradictory knowledge and belief ascriptions. Each one of these philosophical topics have been greatly shaped by the framework of epistemic logic, to the point that today they are almost not analyzed without the use of some epistemic logic.

Nevertheless, in later parts of the 20th century, epistemic logic only played a marginal role in analytic epistemology, possibly due to its apparent incompatibility with the (then prominent) project of reductively analyzing the concept of knowledge. After all, in most applications of epistemic logic, knowledge is taken as a primitive concept (much like in knowledge-first epistemology), an assumption that is in tension with the view that knowledge is a composite, non-primitive, concept (prevalent in late 20th century epistemology). Meanwhile, Hintikka's seminal work in epistemic logic found applications outside of traditional philosophy, including in linguistics and formal semantics (Janssen and Zimmermann 2021), computer science (Fagin et al. 1995) and economics and game theory (Aumann 1999). In the last 20 years, with the influence of knowledge-first epistemology on the one hand (Williamson 2000), and the increasing interaction between formal epistemology and non formal epistemology (such as in Bayesian epistemology) on the other hand (Weisberg 2021), the intersection between epistemic logic and epistemology seems to be rising.

The formal language of propositional, single agent epistemic logic includes propositional logic together with the K propositional operator. More formally, the syntax can be expressed as follows:

$$\varphi := \ p \mid \varphi \wedge \psi \mid \neg\varphi \mid K\varphi$$

where other Boolean connectives are defined as usual. $K\varphi$ is read as “The agent knows that φ.” To bypass certain assumptions involving cognitive psychology, $K\varphi$ is also often read as “The agent is in a position to know that φ." The exact meaning of *being in a position to know* is often left implicit (Williamson 2000: 95, cf. Berto and Hawke 2018). With an emphasis on knowledge of the external world, saying that an agent is in a position to know φ, implies that the agent actually has the evidence or information about the external environment needed to know φ. Under this reading of K, it is acceptable to accept for any tautology φ that $K\varphi$: since knowing tautologies does not require any evidence or information about the external environment, we can accept that in this sense, the agent is in a position to know tautologies. Of course, actual epistemic agents do not know every tautology. Many tautologies have not crossed the mind of actual agents, and some kind of awareness seems to be necessary for actual knowing. However, in this work, epistemic logic is used to track the information the agent has about the external environment, not to distinguish different levels of awareness, nor to study the nature of logical or mathematical knowledge. Therefore, I will switch between the two different readings of K throughout this dissertation. I use $\hat{K}$ throughout the dissertation to denote the dual operator of K, and as usual, treat $\hat{K}\varphi$ as an abbreviation of $\neg K \neg \varphi$. $\hat{K}\varphi$ is read as "The agent cannot rule out φ."

For the semantic framework of epistemic logic, we just give an epistemic interpretation to the standard Kripke, or possible worlds semantics of modal logic (see Blackburn et al. 2002, van Benthem 2010, for a modern introduction). An epistemic model M is a triplet (W, R, V), where is W is a set of epistemically possible states (or worlds), R is an indistinguishability relation, such that Rwu means that the epistemic agent cannot distinguish state w from state u, and V is a valuation function assigning a truth value for every atomic propositional letter at every possible state. The sentence $K\varphi$ is true at a given state w of a model M just in case every state accessible with the R relation satisfies φ. Formally:

$$M, w \models K\varphi \quad \Leftrightarrow \quad \forall u, Rwu : M, u \models \varphi$$

The semantic clause of the dual operator is therefore:

$$M, w \models \hat{K}\varphi \quad \Leftrightarrow \quad \exists u, Rwu : M, u \models \varphi$$

The semantics of the Boolean connectives is the one from propositional logic.

As to restrictions on the R relation, it is agreed upon that the R relation is reflexive: this guarantees, semantically, that the agent can never rule out the actual state of affairs: whatever is actually the case is epistemically possible. Syntactically, this guarantees the factivity of knowledge: the formula $K\varphi \to \varphi$ is true in any reflexive Kripke frame. Further restrictions on the R relation correspond to the introspective abilities of the agent and remain controversial within epistemology (see the subsection Introspection Principles).

1.4.1 Misconceptions about epistemic logic

There are a few common confusions about the semantics of epistemic logic, which the terms possible worlds, indistinguishability relation, and epistemic model often invoke. First, the possible worlds terminology does not carry any commitments about the metaphysical nature of such worlds. Talk about possible states, or possible worlds, is just a way of talking about possibilities. epistemically possible states are states of affairs that, relative to the information of a given agent, cannot be ruled out. When I don't know where my keys are, and I haven't checked the bedroom or the kitchen yet, then I cannot rule out that the keys are in the kitchen, nor can I rule out that they are in the living room. It is possible that the keys are in one place or the other. For simplicity, we reify these possibilities by talking about a possible world in which the keys are in the kitchen, and another possible world in which the keys are in the living room. Assumptions about possible worlds that rise from metaphysics and philosophy of language (e.g. Kripke 1980) do not automatically transfer to epistemically possible worlds.

Second, talk of indistinguishability relation across different possibilities can mistakenly imply an elaborate mental, cognitive, or phenomenological capacities by the agent, or incorrectly interpreted as offering some psychological picture of the agent's inner mental state. Such assumptions are not built into the semantics of epistemic logic, and in fact the semantics of epistemic logic allow for the most naturalistic and behaviouristic interpretations of knowledge ascriptions, including ascription to systems that do not include any mental capacities (like computers and digital sensor) or to non-human animals. A dog can smell a hotdog in the apartment (the dog knows that there is a hotdog in the apartment); the dog checks the kitchen and living room (so the dog does not know the exact location of the hotdog). Using the conceptual resources available to us from the semantics of epistemic logic, we say that the dog cannot distinguish between the possibility in which the hotdog is in the kitchen (one possible world) from the one in which it is in the living room (a different possible world): the two worlds are indistinguishable, given the dog's epistemic state (which is itself extracted from the dog's behaviour, in this example). Of course, the semantics of epistemic logic does not force upon a behaviouristic or naturalistic interpretation; the semantic framework is relatively neutral.

Third, epistemic models are *models*. As such, their purpose is to represent an aspect of reality, at the expense of other aspects. An epistemic model represents an epistemic state; it is not identical to that state, nor does it reduce that state into something else. Compare: the New York subway map is a model of the New York subway systems, highlighting some of the aspects of the system (the order of the stations on a given line), on the expense of other aspects (the metric distance between the stations, or the exact location of the subway tracks). Disregarding an epistemic model just on the accusation that it is 'unrealistic' can be as imprudent as not consulting the New York subway map as it does not offer a realistic depiction of New York city. Like any model, an epistemic model is to be judged (as accurate, or realistic) relative to its purpose. In this dissertation, the purpose of using epistemic models is to represent reasoning about changes of states of knowledge. The models offer a convenient map for tracking the agents' knowledge about their own change of knowledge; they are very useful in that regard. Their purpose is not to capture every possible aspect of human knowledge.

1.4.2 Axioms of epistemic logic and their interpretation

Candidate axioms of systems of epistemic logic include:
K: $K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$
T: $K\varphi \rightarrow \varphi$
4: $K\varphi \rightarrow KK\varphi$
5:$\neg K\varphi \rightarrow K\neg K\varphi$.

The last two candidate axioms are considered as introspection principles, as they stipulate what the agent knows about their own knowledge. Axiom **4** is also known in epistemology as the **KK principle**, **positive knowledge introspection**, or just as **positive introspection**; axiom 5 is known as **negative knowledge introspection**, or just as **negative introspection**. The **T** axiom is also known as **factivity**, as it captures the fact that knowledge is factive. The **K** axiom is known as a **closure**, axioms as it captures the assumption that knowledge (or being in a position to know) is closed under entailments.

The logic generated by the addition of **K** and **T** to the axioms of propositional logic (together with the necessitation rule: $\vdash \varphi \Rightarrow \vdash K\varphi$) is known as the logic **T**. Adding axiom **4** to system **T** results in system **S4**; adding axiom **5** to system **S4** (or just to system **T**) results in system **S5**. Therefore, discussions on whether **T**, or **S4**, or **S5** are the right system to represent knowledge usually just amount to discussions about whether to accept or reject a particular introspection principle. It should be noted that there are infinitely many modal systems between **S4** and **S5** alone, each one with a particular set of introspection principles (Chagrov and Zakharyashchev 1992), so the question of the 'true' epistemic logic goes beyond the two axioms **4** and **5**. Moreover, the search for the true logic of knowledge is somewhat like the search for the true map of New-York city: the accuracy of the representation system cannot be judged without considering its purpose.

The logic **S5** is complete with respect to all Kripke frames where the R relation is an equivalent relation; **S4** is complete with respect to all reflexive transitive Kripke frames; **T** is complete with respect to all reflexive frames (see Blackburn et al. 2002). Throughout the dissertation, I will call a Kripke model with an equivalence relation an **S5** model, and similarly for **T** and **S4** models.

The axioms of epistemic logic are often assumed to describe ideal epistemic agents. Such an assumption misses an important distinction between the axioms. Some epistemic principles describe an epistemic *ability* an ideal agent might have; others aim to capture some *metaphysical* property of knowledge; independent of abilities. The **K** axiom, when $K\varphi$ is read as "the agent knows φ", assumes that agents are logically omniscient, and therefore only applies to ideal agents. It is less clear whether the *in a position to know* reading of $K\varphi$ implies logical omniscience, given the **K** axiom (See Stalnaker 1991). In any case, it is important to contrast the ideal nature of the **K** axiom against the descriptive nature of the **T** axiom. The truth of the **T** axiom has nothing to do with the abilities of ideal agents; it follows from metaphysical assumptions about the knowledge state of any agent (ideal or not). Therefore, it is mistaken to assume that any true formula of epistemic logic only applies to ideal agents, and when using epistemic logic, one has to be attuned to this difference (the object language of epistemic logic does not offer any tool to draw the line between formulas being true in virtue of an idealized *ability* assumption, and those being true based on *metaphysical* assumptions about knowledge).

This difference is often ignored when philosophers debate the truth of introspection principles, like **4**. Under an ideal ability reading, **4** says that if an ideal agent is in a position to know φ, then such agent is able to know that they know φ. Under such reading, **4** describes an *epistemic ability* that we assume that ideal agents have (if we accept the axiom, much like **K**); we can still agree that **4** fails for non-ideal agents. One can also read **4** in non ideal, descriptive manner, stating a metaphysical fact about knowledge. Under such reading, **4** states that the knowledge state is *idempotent*: the epistemic state of being in a position to know φ is just *identical* to the state of being in a position to know that one is in a position to know φ. In other words, that it is metaphysically impossible to be in an evidential state that makes the formula $K\varphi$ true and $KK\varphi$ false, or that the formulas $K\varphi$ and $KK\varphi$ just refer to the same epistemic state. Indeed, even classifying the **4** axiom as an introspective principle is misleading from the perspective of the idempotency reading of the axiom, as talk of 'introspection' can invoke the image of an activity or ability on the part of the agent (the *ability to introspect*). The ideal, ability reading of **4** is quite different than the metaphysical, idempotency reading.

1.4.3 Topological semantics for epistemic logic

Beyond the standard Kripke semantics for epistemic logic, there exists a prominent alternative *topological* semantic framework for the modal logic **S4**, with an interesting epistemic interpretation. Historically, the topological semantics for **S4** preceded Kripke's possible worlds framework (McKinsey and Tarski 1944). This subsection offers a very brief introduction to this topic, which is further developed in Section 2.4 of Chapter 2. For a proper presentation of topological semantics for modal logic, see e.g. van Benthem and Bezhanishvili (2007) and Baltag et al. (2019).

Given a set X, the pair (X, τ) is called a *topological space*, where τ is a collection of subsets of X that satisfy the following properties: $\emptyset$ and X are in τ; every finite intersection of sets in τ is in τ; and every arbitrary union of sets in τ is in τ. The members of τ are called *open sets*. An open set X containing x is called an *open neighborhood* of x. Given a topological space (X, τ), a topological modal model (X, τ, V) is a topological space together with a valuation function V that assigns a truth value for every propositional formula for every point x in X (like the valuation function in Kripke semantics).

An epistemic formula $K\varphi$ is true in a given point x of a topological model if there exists an open set U containing x in which φ is true in every point in U. In simpler words, $K\varphi$ is true in x if x is part of a φ open set. Formally, this can be stated as

$$M, x \models K\varphi \quad \Leftrightarrow \quad \exists U, U \in \tau : (x \in U \ \ \& \ \ \forall y \in U : M, y \models \varphi)$$

The duality between $K\varphi$ and $\hat{K}\varphi$ implies the following truth condition for $\hat{K}\varphi$:

$$M, x \models \hat{K}\varphi \quad \Leftrightarrow \quad \forall U, U \in \tau : (x \in U \quad \Rightarrow \quad \exists y \in U : M, y \models \varphi)$$

The remaining truth clauses are the same as in Kripke semantics. One can check that the topological semantics validate $K\varphi \to \varphi$ and $K\varphi \to KK\varphi$ (the **T** and **4** axioms). In fact, the topological semantics are sound and complete with respect to the modal system **S4**.

Although the above semantics might look foreign to philosophers accustomed to Kripke's possible worlds semantics approach to modal logic, the semantics has a very simple epistemic interpretation (arguably even simpler than the Kripke semantics interpretation). Given a set X of epistemic possibilities (or possible worlds), one can intuitively think of the open sets in τ as the *pieces of evidence* or *pieces of information* available to the agent in a given state (see e.g. Baltag et al. 2019: 221). If, at a given state x, the agent has a piece of evidence for φ, then the agent knows φ. The topological semantic clause for $K\varphi$ says just that. If the agent has no piece of evidence that distinguishes between state x and state y (at a given point), then these states are epistemically indistinguishable to the agent (at that point).

Under this interpretation, the topological definition of an open set puts epistemological constraints on the notion of evidence: for one thing, this interpretation makes evidence to be factive (much like in the externalist view on evidence). As such, the topological interpretation can be, and has been, criticized as a semantics for evidence and knowledge and loosened to allow for different approaches to evidence. Recent work on *evidence logic* (van Benthem et al. 2014) uses *neighborhood semantics* (another alternative to Kripke semantics, which generalizes both the latter and topological semantics) to logically study a much looser notion of evidence. In such a framework, not only can an agent have false pieces of evidence, but pieces of evidence can contradict each other. From a more epistemological perspective, Williamson (2000: 125) criticises and rejects a topological interpretation of epistemic safety (a condition necessary for knowledge, according to him). Like Kripke semantics, some of the possible critiques of the topological semantics can arise from confusing epistemic *states* with epistemic *models*, and from reading various cognitive or psychological assumptions into the formal framework. The comments I have addressed in Section 1.4.1 apply to topological semantics as well.

The connection between topology and epistemology is broader. Topological semantics have been also used to study the notion of full belief (Baltag et al. 2019, see Section 1.5 for more on the notion of full belief) and the dynamics of knowledge and belief (Baltag et al. 2015, Bjorndahl 2018, see next subsection). Epistemic interpretations of topological structures play an important role in learning theory as well (Kelly 1996).

1.4.4 Dynamic epistemic logic

There are many ways to extend the basic logic presented above. *Dynamic epistemic logic* (DEL) is an extension of epistemic logic which allows to reason about change of knowledge due to the reception of new information. Chapter 2 offers a more extensive introduction to DEL. In DEL, a learning event, or update operator $[!\varphi]$ is added to the language of epistemic logic, such that $[!\varphi]\psi$ is read "as a result of updating with φ, ψ is the case." By treating learning events as propositional operators, DEL allows both the modeler and the modeled agent to reason about the effects of updates. The scoping patterns $[!\varphi]K\psi$ represents the a-posteriori, or resulting, epistemic state of an agent (as a result of updating with φ, the agent is in a position to know ψ); the scoping patterns $K[!\varphi]\psi$ represents the agent a-priori, or initial knowledge state about the effects of an update (the agent is in a position to know that as a result of updating with φ, ψ is the case). Here, the terms *a-priori* and *a-posteriori* are used relative to some learning event (i.e. update), not in an absolute sense. This flexibility distinguishes DEL from other popular frameworks for representing learning in epistemology, such as Bayesian epistemology and AGM belief revision, which offer simple ways of representing the agent's change of epistemic state given new information, but not necessarily a way of representing the agent's information about that change of information.

One commonality between frameworks like standard DEL and Bayesian epistemology is that learning events are *individuated* using the informational content of the learning event. In simple DEL, every learning event $[!\varphi]$ is equated with an update with some φ of the object language; in Bayesian epistemology, every posterior state $P_A(B)$ is equated with a prior state $P(B|A)$, the result of conditioning on some sentence A of the object language. In essence, this assumes that the content of every possible learning event is a-priori transparent to the agent: every event is just learning a concrete piece of information φ.[1] This assumption rules out situations in which the agent is uncertain about the content and the exact effect of the learning event. In Chapter 2 of this dissertation (and, to a lesser extent in Chapter 5), I develop a more general version of dynamic epistemic logic in which the content and effect of learning events are not transparent to the agent. I use this logical framework in various philosophical debates in the dissertation. Throughout the dissertation, I argue that opaque (i.e. non-transparent) updates, and the distinct type of ignorance they bring about, should play a more central role in epistemology, especially in an externalist one.

[1] Exception to this observation include Jeffery conditioning, in which the learning event is not necessarily made explicit as a sentence of the language (Jeffrey 1965), and *event models* in DEL (Baltag and Renne 2016).

1.5 Introspection principles

Within epistemology, debates about introspection often combine the use of some epistemic logic. In this section, I survey some important epistemological arguments concerning introspection, and their connection to epistemic logic.

There is a simple and powerful argument against an unrestricted form of axiom **5**, negative knowledge introspection, $\neg K\varphi \to K\neg K\varphi$, that involves the interaction between knowledge and full belief. By full belief, we mean a doxastic state that is indistinguishable from knowledge, from the agent's perspective. Suppose that the agent falsely believes p (in the sense of full belief): $Bp \wedge \neg p$. Since p is false, it cannot be known, due to the factivity of knowledge: $\neg p \to \neg Kp$, therefore, $\neg Kp$. According to negative introspection, it follows that the agent knows that they do not know $p, K\neg Kp$. In this situation the agent fully believes p and knows that they do not know p. Now, it seems that we have reached a contradiction: if full belief is indistinguishable from knowledge, then, since the agent fully believes p, for all the agent knows, they know p ($\hat{K}Kp$). At the same time, due to negative introspection, the agent knows that they do not know p. This is a contradiction.

A similar, but weaker, argument against negative introspection can be made without an appeal to the notion of full belief. Combining **T**, $\neg\varphi \to \neg K\varphi$ (in its contra-positive form) with **5**, $\neg K\varphi \to K\neg K\varphi$ implies the principle $\neg\varphi \to K\neg K\varphi$ (by the transitivity of $\to$): if a φ is false, then the agent knows that they don't know φ. Nothing in our conception of knowledge suggest such a strong assumption. As a result, no contemporary epistemological view endorses axiom **5** in an unrestricted fashion.

The connection between full belief and knowledge has proven to be quite closer, given the right introspective assumptions. It turns out that under relatively weak epistemic assumptions, belief can be defined in purely epistemic terms, as the epistemic possibility of knowledge (Lentzen 1978, Stalnaker 2006). The epistemic-doxastic principles that are assumed in such framework are the following:

$K\varphi \to B\varphi$ **Knowledge entails belief**

$\neg B\varphi \to K\neg B\varphi$ **Negative belief introspection**

$B\varphi \to \hat{K}K\varphi$ **Full belief**

The first principle states that knowledge entails belief, and it is widely accepted within epistemology (Ichikawa and Steup 2018). The second principle is a weak version of negative introspection, stating that if the agent does not fully believe φ, then they know that they do not believe φ. This weak form of negative introspection is not susceptible to the argument presented earlier against the **5** axiom. Since belief is not factive, the principle $\neg\varphi \to \neg B\varphi$ does not hold, which is needed in the argument against **5**. The third principle, **full belief**, follows from the assumption that the agent cannot distinguish full belief from true knowledge. Such assumption implies that if an agent fully believes φ, the agent cannot rule out that they know φ (if they could, then the agent could distinguish the belief state from the real knowledge state). Hence $B\varphi \to \hat{K}K\varphi$.

The claim that full belief just amounts to the epistemic possibility of knowledge is equivalent to the principle $B\varphi \leftrightarrow \hat{K}K\varphi$. The left-to-right direction is just the **full belief** assumption. To establish the right-to-left direction of the principle, we can just prove the statement $\neg B\varphi \to K\neg K\varphi$ (the contrapositive form). From $\neg B\varphi$, $K\neg B\varphi$ follows by **negative belief introspection**. Since we take $K\varphi \to B\varphi$ as a validity, $K(K\varphi \to B\varphi)$ follows by necessitation. The latter is equivalent to $K(\neg B\varphi \to \neg K\varphi)$. Given $K\neg B\varphi$ and the closure of knowledge under entailment (axiom **K**), $K\neg K\varphi$ follows. This reasoning establishes the equivalence $B\varphi \leftrightarrow \hat{K}K\varphi$ from the above doxastic-epistemic principles.

This logical reduction of belief in terms of knowledge aligns with knowledge-first epistemology: it shows how to treat belief as a derivative notion of knowledge. It is related to the idea that full believing can be analyzed as *acting as if you know* (Williamson 2000). It is interesting to note that while the epistemological attempt to reduce knowledge in terms of belief has resulted in more and more complex epistemological epicycles (in the form of different approaches to justification, see Ichikawa and Steup 2018) without ever reaching a consensus, the opposite attempt of reducing belief in terms of knowledge is as straightforward as the acceptance of two or three epistemic logic principles, and *without* the addition of any further epistemological concepts, beyond knowledge or belief. Nevertheless, within philosophy as a whole, understanding belief in terms of knowledge remains relatively unpopular, with a few, but notable, defenders (e.g. Stalnaker 2006, Williamson 2000).

Within epistemic logic, however, the $B\varphi \leftrightarrow \hat{K}K\varphi$ conception of belief has become quite dominant. Note that $B\varphi \leftrightarrow \hat{K}K\varphi$, together with **5** implies that $B\varphi \leftrightarrow K\varphi$. For $\neg B\varphi \leftrightarrow \neg K\varphi$ can be replaced with $K\neg K\varphi \leftrightarrow \neg K\varphi$ (given the definition of belief in terms of knowledge); the left-to-right direction follows from factivity, the other direction follows directly from **5**. Therefore, in contemporary logics of knowledge and belief the logic of the K operator is situated below **S5** and above **S4**. Such epistemic-doxastic logics have proven to be immensely successful in studying the dynamics of knowledge update and belief revision, and related epistemic notions such as subjective probabilities and the connection to topology (van Benthem 2012, Baltag and Smets 2006, Baltag et al. 2019).

Only a very extreme form of access internalism about knowledge and justification can accept a principle like **5**, since, as the last paragraphs have shown, such principle implies that the agent can introspectively distinguish between knowledge and false beliefs (and so align one's knowledge with one's beliefs). The radical skeptic about the external world can also be seen as accepting **S5**: the skeptic vacuously matches belief to knowledge (by having no knowledge, and suspending all beliefs).

The status of the **4** axiom (the KK principle) is much more controversial within epistemology compared to that of **5**. The principle has its contemporary defenders within epistemology (Das and Salow (2016), Dorst (2019), Goodman and Salow (2018), Greco (2014a, 2014b, 2015a, 2017), Stalnaker (2015)), but it is fair to say that defending the KK principle is a minority view.

It is often assumed that externalism is in tension with the KK principle (see, e.g., Okasha 2013).

The general line of thinking is that externalists, unlike internalists, do not presuppose that epistemic agents have the introspective ability to access the justification for their knowledge. Therefore, it is natural to assume that externalist agents do not always have the ability to be in a position to know that they know. According to a related argument, accepting the KK principle makes it harder to know. In order to know φ, one must also know $K\varphi$, and $KK\varphi$, etc, *ad infinitum*, leading to a regress problem. Externalists, as well as naturalists and many other epistemological theories, want to allow for less sophisticated epistemic agents (like non-human animals) to have knowledge, and it is questionable whether such agents even have the epistemic or cognitive capacity to have higher order knowledge. The common assumption of such arguments is that having higher-order knowledge is epistemically harder than having first-order knowledge, and that therefore, there is no reason to require agents such a demanding standard. See (Greco 2015b) for a survey of such arguments.

The weakness of these arguments is that they attack the *ability* reading of the **4** axiom (the KK principle), while ignoring the *metaphysical* reading. Recall that according to the ability reading, the KK principle attributes an epistemic ability to epistemic agents: the ability to come to know that one knows. According to the metaphysical reading, the knowledge state is just idempotent: the sentence $K\varphi$ and $KK\varphi$ just refer to the same epistemic state. Any argument that presupposes that attaining higher order knowledge is epistemically harder than first order knowledge implicitly rejects the metaphysical reading of the KK principle. If $K\varphi$ and $KK\varphi$ just refer to the same epistemic state, then it is senseless to say that being in one state is harder than being in the other. Likewise, if the sentence *the dog is in a position to know that there is a hotdog in the apartment* just denotes the same situation as *the dog is in a position to know that the dog is in a position to know that there is a hotdog in the apartment*, then the cognitive capabilities of the dog are irrelevant to the truth of the KK principle. On the contrary, from an externalist, behaviouristic perspective one should ask: what kind of observable situation could even distinguish the one description from the other? Recall that the externalist would *not* wish to appeal to the agent's inner psychological states to determine knowledge, so considerations regarding the dog cognitive capacities become largely irrelevant for the externalist. In that sense, externalism about knowledge is very much compatible with the idempotent reading of **4**.

In the same way, the above mentioned regress argument against the KK principle does not apply to the metaphysical reading, as $K\varphi, KK\varphi$ and $KKKKKKK\varphi$ just all refer to the same state under the latter reading. The absolute value function, $abs(x)$, is idempotent. Someone who fails to note this fact might think that $abs(-6), abs(abs(-6))$ and $abs(abs(abs(abs(abs(abs(-6))))))$ refer to different numbers. Since the iteration of the absolute value has no effect, these are just different ways to denote 6. The same confusion afflicts those who use the regress argument against the idempotency (or metaphysical) reading of the KK principle. The regress argument is only applicable against those who reject the metaphysical reading of the principle, but accept the ability reading.

The most influential argument against the KK principle in contemporary epistemology is arguably Williamson's *inexact knowledge*, or *margin-for-error* argument (Williamson 2000, 2014). Williamson's argument is analyzed and criticized in detail in Chapter 5 of the dissertation. Williamson's argument clearly rejects both the idempotency and the ability readings of the KK principle. He offers a direct analysis of the different pieces of evidence needed in order to attain different levels of higher-order knowledge, in direct opposition to the idempotency reading of the knowledge operator.

The details of the argument are made explicit later in the dissertation. Briefly, the argument uses epistemic logic to offer a counter model to the KK principle. Intransitive epistemic Kripke models invalidate the **4** axiom. Williamson argues that the epistemic state of inexact knowledge should be best characterized as one in which the agent's indistinguishability relation is intransitive. To argue for that, Williamson formulates an epistemic margin-for-error principle that must hold in cases of inexact observational knowledge. The intransitivity of inexact knowledge invalidates the **4** axioms and offers a counter example to the KK principle. In Chapter 5, I argue that Williamson confuses the epistemic event of making an inexact observation with the epistemic state resulting from such observation. Inexact observations are indeed non-transparent, and conflict with the introspective abilities of the agent. But the tension is with a dynamic form of introspection, not with the KK principle. I introduce the notion of dynamic introspection in the next section.

1.6 Dynamic Introspection

I use the term *dynamic introspection* to contrast it with (what I call) static introspection. Static introspection is just the standard introspection principles from static (i.e., non-dynamic) epistemic logic, like the **4** and **5** axioms discussed above. I show in Chapter 2 that the two notions of introspection are logically independent.

To articulate principles of dynamic introspection, let us consider again the model of knowledge of the external world I started this chapter with. The model had three elements: the initial epistemic state, the learning event from the information source, and the resulting epistemic state. It is depicted in Figure 1.1.

Learning event

Initial epistemic state ⟶ **Resulting epistemic state**

Figure 1.1: The simple learning model

This model is deceptive because (among other issues) it allows us to express the ignorance the agent has at the different stages of the process (the initial state vs. the resulting state), but it makes it harder to think of the ignorance the agent has about the process itself. In Figure 1.1, uncertainty is seen as something that applies at particular stages of the process; but what about uncertainty about the process as a whole? If we just stick with the simple model, we might end up ignoring the uncertainty the agent has about the dynamic process *itself*.

What the agent knows about the dynamic process itself needs to be divided into what they know at the initial state and what they know in the resulting stage. From the perspective of initial state, the agent might not know what are the epistemic effects of a learning event. The agent might not be able able to answer the forward looking question: *what do I know about the effects of the incoming learning event?* Figure 1.2 complicates the simple model to accommodate this type of dynamic uncertainty.

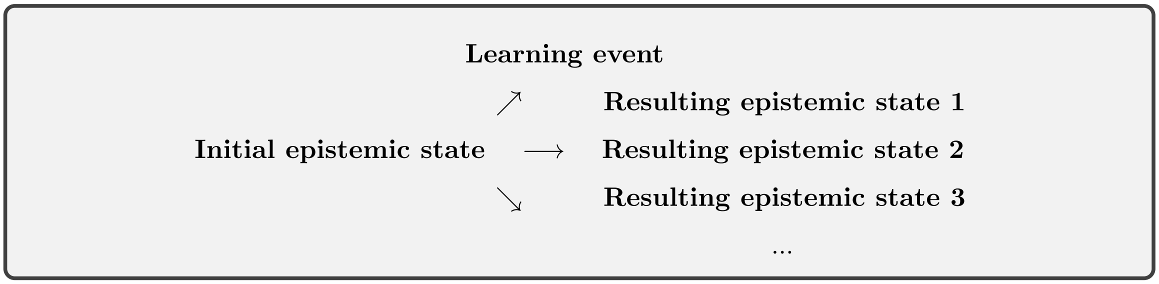
Learning event
Initial epistemic state
Resulting epistemic state 1
Resulting epistemic state 2
Resulting epistemic state 3
...

Figure 1.2: Forward looking uncertainty

In Figure 1.2, in the initial epistemic state, the agent is uncertain as to how the (fixed) learning event is going to effect them: they don't know if the learning event results in epistemic state 1, or 2, or 3, etc.. The simple model in Figure 1.1 does not allow for a direct representation of this type of uncertainty.

From the resulting epistemic state, the agent needs to consider the backwards looking question: *what epistemic event brought me to my current situation?* This type of uncertainty is depicted in Figure 1.3.

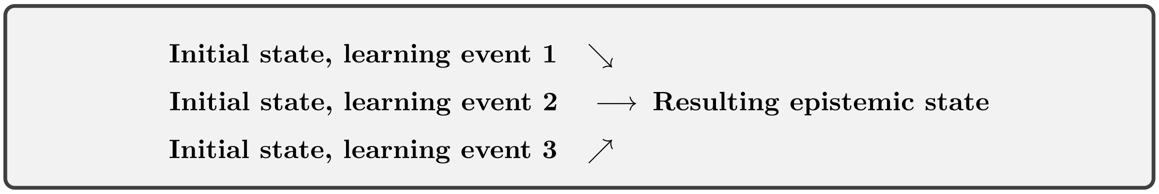
Initial state, learning event 1
Initial state, learning event 2
Initial state, learning event 3
Resulting epistemic state

Figure 1.3: Backward looking uncertainty

In Figure 1.3, in the resulting epistemic state, the agent is uncertain what learning event brought them to their current state (was is learning event 1? or 2? etc.).

Dynamic introspection principles answer these questions in a principled manner. They specify what the agent knows about their own learning events: unlike static introspection, which specifies knowledge of knowledge in a principled manner, dynamic introspection specifies knowledge of learning events. And like static introspection principles, dynamic introspection principles can fail, leaving the agent with ignorance about the above mentioned forward and backward looking questions.

To make these question less vague let's consider the simple case of looking at a clock and coming to know the time. Suppose that prior to looking at the clock (in the initial epistemic state) the agent does not know that the clock is reliable. Therefore, the agent does not know that looking at a clock results in coming to know the right time. The agent has forward looking uncertainty (the type depicted in Figure 1.2): they do not know whether the event of looking at the clock results in an epistemic state in which they know the time (if the clock is reliable) or an epistemic state in which they do not know they time. Recall that the skeptic argues that since the agent does not know that the clock (the source of information) is reliable, the agent does not come to know the time by looking at the clock (the skeptical conclusion). Here, the skeptic is committed to the following dynamic introspective principle: *in order to come to know something from a source of information, the agent has to know that the source of information accurately represents reality.* The skeptic uses this principle in its contrapositive form: since the agent does not know that the source of information accurately represents reality, the agent does not come to know the information from the source. This dynamic introspective principle is known as the **no-miracles** principle (the originates from epistemic game theory, see van Benthem and Klein 2018). It concerns the forward looking question we mentioned before, it describes an *ability* the agent has, and it states that if the agent gains knowledge from a source of information at the resulting stage, then the agent must have known, at the initial stage, that the source of information is reliable (i.e. results in knowledge). Although this epistemic principle might seem overly demanding, I argue in the dissertation that it is implicitly assumed in important debates in epistemology, such as the problem of perception. This is evident by the fact that formal frameworks for representing updates (including DEL and Bayesian epistemology) implicitly assume the **no-miracles** principle, and the fact that, to the best of my knowledge, there exists no epistemological literature on that principle. In Chapter 2, I present a logical framework that invalidates the no-miracles principle. In Chapter 3, I argue that the principle plays a central, albeit implicitly so, role in epistemology, and I advocate for an epistemological picture that explicitly rejects **no-miracles**.

The **no-miracles** principle concerns what the agent has to know about the source *prior* to the learning event in order to gain knowledge from it. A dual dynamic introspection principle, known as **perfect-recall** concerns what the agent knows about the epistemic event that resulted in the agent current (resulting) epistemic state. More accurately, the principle states that *if the agent knows a certain result of a learning event prior to its occurrence (in the initial stage), then the agent comes to know that this result holds posterior to the event (at the resulting stage).* Example: suppose that in an initial stage, the agent knows that if they come to know the time by looking at the clock, then they know that they just looked at a clock. According to **perfect-recall**, this implies that when the agent looks at the clock and come to know the time, they also come to know that they just looked at a clock. As the name suggests, **perfect-recall** can fail if the agent loses memory. Suppose that, for some reason, right after looking at the clock, the agent's memory of looking at the clock (but not of the time) is being erased. The agent knows the time, but they forgot how they know it. In that scenario, in the resulting epistemic state the agent does not know that they just looked at a clock (as they do not remember actually looking at the clock). The agent does not know what epistemic event brought them to their current (resulting) epistemic state, much like the situation depicted in Figure 1.3. Such a scenario violates the **perfect-recall** principle.

Although **perfect-recall** has been discussed in formal epistemology, especially in the context of memory loss (like the sleeping beauty problem, see Halpern (2004)), I argue in chapters 2 and 4 of the dissertation that the role of **perfect-recall** is much more central in epistemology, and its interpretation goes much further than memory loss. Essentially, whenever the agent knows something, but does not know how they know it, the failure of perfect-recall is involved. Like **no-miracles**, I argue that **perfect-recall** is implicitly assumed in the bootstrapping problem, partially due to the fact that many presentations of the bootstrapping problem do not have the resources to explicitly formulate **perfect-recall**, and its failure.

In Chapter 2, I show how to logically formalize the type of uncertainty that is depicted in Figures 1.2 and 1.3. Throughout the dissertation, I develop an epistemological picture that explicitly rejects the **no-miracles** and **perfect-recall** principles. Epistemic agents can gain knowledge from a source of information, without knowing almost anything about it. The ignorance about the source of information (and the opacity of the corresponding learning event) is reflected in the rejection of these dynamic introspection principles. The combination of knowledge of the external world together with dynamic ignorance about the process of gaining information results in a novel type of externalist epistemology.

Chapter 2

Opaque updates

Chapter abstract: If updating with E has the same result across all epistemically possible worlds, then the agent has no uncertainty as to the behavior of the update, and we may call it a transparent update. If an agent is uncertain about the behavior of an update, we may call it opaque. In order to model the uncertainty an agent has about the result of an update, the same update must behave differently across different possible worlds. In this chapter, I study opaque updates using a simple system of dynamic epistemic logic suitably modified for that purpose. The chapter highlights the connection between opaque updates and the dynamic-epistemic principles Perfect-Recall and No-Miracles. I argue that opaque updates are central to contemporary discussions in epistemology, in particular to externalist theories of knowledge and to the related problem of epistemic bootstrapping, or easy knowledge. Opaque updates allow us to explicitly investigate a dynamic (or diachronic) form of uncertainty, using simple and precise logical tools.

2.1 Introduction

There is a widespread notion of *update* in formal epistemology that can be semantically summarized in Figure 2.1.

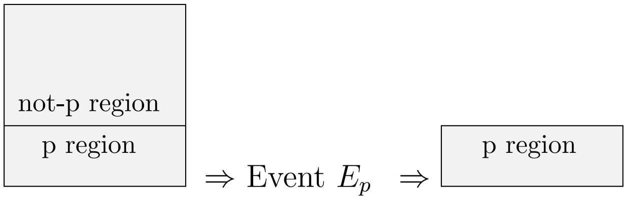

Figure 2.1: Update as a world insensitive function.

We have some prior, or initial, model on the left-hand of the Figure, containing both p and not-p worlds, representing a situation in which an agent is uncertain whether p is the case. The event E of receiving the information that p results in a new model in which the not-p worlds are eliminated (the model to the right), a model representing a situation in which there is no uncertainty as to p. This simple sketch of updating is at the basis of many systems that represent information change, including the Stalnakerian notion of assertion, Dynamic Semantics, Bayesian updating, and various dynamic epistemic logics.

Of course, each implementation of this basic skeleton idea is different, but here I want to point our attention to an assumption that can be detected even at this level of abstraction. The picture in Figure 2.1 portrays the event of learning p as a transition from one model to the other, a model transformer. In other words, the event of receiving the information that p is understood as a function (sometimes partial) from epistemic models to epistemic models. Since such a function is assumed to send us from one *model* to another, it is *insensitive* to the world of evaluation. In other words, the function behaves the same at every possible world of the prior model—at each world, the function sends us to the same posterior model. Put differently again, we don't need to know which world is considered actual in order to compute the model which results from an update with p. At the same time, the picture in Figure 2.1 builds on the idea that when something is the same in all possible worlds, there is no uncertainty about it. Putting these two threads together, since the picture in Figure 2.1 assumes that the update behaves the same in each possible world, and since certainty is assumed to be agreement across all possible worlds, we essentially assume that *the agent has no uncertainty as to the behavior of the update.* Thus, the update in Figure 2.1 is in some sense *transparent* to the agent.

In this chapter, I analyze updates that are not transparent, but *opaque.* If we want to represent the uncertainty the agent has about the effect of an update, it should be possible to have situations in which the same update behaves differently in different epistemically possible worlds. Such situations are not just meant to generalize the notion of update for purely technical reasons. The second main theme of my chapter is that modeling opaque updates is, as I will argue, quite relevant to various debates in contemporary epistemology, and in particular, to the broad position of *externalism.*

For instance, a reliabilist about knowledge argues that the effect of the same learning event can be different in a world in which the source of information is reliable as opposed to a world in which the source is unreliable, while assuming that the agent does not know if the source is in fact reliable. Consider a simple scenario: you look at a clock tower that has, in fact, a reliable clock mechanism. You don't know, however, that the clock mechanism is reliable. According to reliabilists (and many other externalists about knowledge and evidence), it is possible for you to come to know the time by looking at the clock, even though you don't know the clock is reliable. Thus, prior to looking at the clock, you cannot know whether looking at the clock will result in a situation in which you know the time (if the clock is in fact reliable) or in a situation in which you don't know the time (if it is unreliable). Externalists think of the event of looking at the clock as *opaque*: the agent is uncertain as to the epistemic result of the event.

In this chapter, we will see how to model such situations in a simple possible-worlds framework. The formal idea is to build an update U s.t. U sends us to different updated models in different worlds of the prior model. In world w of the prior model in which r is true (the clock is reliable), U will send us to an updated model in which the agent knows the time. In world u of the prior model in which $\neg r$ is true (the clock is unreliable), U will send us to a different updated model in which the agent does not know the time. Since both worlds u and w are initially open to the agent, the agent does not know whether the update with U will result in knowledge of the time or not. We will also explore a backwards form of diachronic uncertainty involving the posterior model: opaque updates can be such that the agent does not know whether it is update U that brought them to their posterior epistemic state, or U' (an update distinct from U).

Externalism about knowledge is roughly the idea that factors external to the epistemic agent can determine the difference between having knowledge and having mere true belief, even if the agent cannot notice those factors from their own (internal) perspective. Reliabilism is just one example. In this chapter, I show how to formalize the externalist intuition when it comes to *epistemic change*. It is possible to construct epistemic updates whose results depend on external factors that the agent is ignorant about. The updates that capture the externalist intuition are opaque updates.

The epistemic logic literature already contains influential endeavors to formalize externalist intuitions. The important works of Rott (2004), Stalnaker (2006) and Baltag and Smets (2008) show how to obtain an externalist conception of knowledge as belief that is stable under revision with true information. However, these approaches do not focus on at least two issue that have become increasingly relevant in the contemporary externalist literature. First, since the above mentioned approaches rely on a 'sphere system' semantics to model belief revision, it follows that the resulting conception of knowledge is positively introspective, i.e. validates the KK principle which states that $K\varphi \to KK\varphi$. Since the vast majority of externalists in epistemology take the *rejection* of the KK principle as essential to externalism, they will reject these formal approaches.[1] Second, the above mentioned formulations seek to analyze knowledge in terms (of some properties) of belief. A prominent camp within externalism, the 'knowledge-first' one, dismisses the project of analyzing knowledge in terms of belief.[2] Therefore, knowledge-first externalists will not accept these approaches. The formulation I propose here can shed new light on these issues: the logic I present for opaque updates is compatible both with the acceptance of the KK principle and with its rejection. Moreover, the logic does not make any assumptions about the relations between knowledge and belief (a belief modality can of course be added to the logic, but I will not explore this here). The model I offer aims to remain as neutral as possible in its epistemological assumptions.

[1] See Okasha (2013) for a survey on the connection between externalism and the rejection of KK.

[2] See Williamson (2000), Nagel (2013), and the collection of chapters in Carter et al. (2017).

Timothy Williamson has also offered models of epistemic logic for studying externalist conceptions of knowledge (2000, 2013, 2014). In these models, the agent's *lack* of positive introspection plays a central role. However, Williamson's models are completely static, remaining silent with respect to the question how externalism construes *change of knowledge.* The dynamic formulation I offer here is meant to bridge this gap.

Within formal epistemology, we are used to model the information the agent has as a set of possible worlds. *Static* epistemic logic has taught us that assuming that such a set is constant across possible worlds amounts to assuming that the agent has *full static* introspection. Philosophical work has concluded that such assumption is highly debatable. Analogously, dynamic epistemic logic teaches us, as I will argue, that assuming that the result of an update is constant across possible worlds amounts to assuming that the agent has *full dynamic* introspection, or full dynamic transparency. Is such assumption justified?

The aim of this chapter is to introduce, analyze and offer a simple working model for an epistemic-logic phenomenon: opaque updates. Rather than advancing an entirely novel contribution to the logical literature, or offering a philosophical argument for or against externalism, my goal in this chapter is to *bridge* logical and epistemological bodies of work. I show how the right application of existing logical tools can be used to offer a fresh approach to well-known problems in epistemology, problems that can benefit from an accessible formal model.

In Section 2.2, I show how to expand a simple dynamic epistemic logic such that it will be able to accommodate opaque updates. Along the way, I highlight and discuss important epistemological assumptions that are built into basic dynamic epistemic logics. The extension uses a familiar combination of propositional dynamic logic with dynamic epistemic logic. Such an extension allows us to violate the basic update axioms of dynamic epistemic logic, the No-Miracles and Perfect-Recall axioms. The combination of these two axioms can be seen as the syntactic analog of the semantic idea from Figure 2.1. In Section 2.3, I argue for the relevance of opaque updates (and the rejection of the No-Miracles and Perfect-Recall axioms) to what is known as *basic knowledge* theories within epistemology, and to the related discussion about the problem of epistemic bootstrapping. Section 2.4 discusses and compares the phenomenon of opaque updates in other logical frameworks. Section 2.5 concludes. Although the technical details of opaque updates are presented for a particular version of dynamic epistemic logic, the broader aim of this chapter is to introduce notions and principles that go beyond any particular system. Whenever applicable, I will suggest connections to the Bayesian framework of updating (these remarks, however, won't be as rigour as the main discussion).[3]

[3] Moreover, since I use epistemic logic, my focus is on knowledge, while the Bayesian framework is meant to model beliefs. Nevertheless, opaque updates can come up in a doxastic setting as well.

2.2 Dynamic epistemic logic with opaque updates

We will use the single-agent *Public Announcement Logic* (PAL) as our base dynamic epistemic logic, which will then be extended to accommodate opaque updates that violate the axioms of PAL. Before we get there, I start with a brief recap of standard PAL. Readers familiar with the basics of PAL can jump to Section 2.2.2.

2.2.1 Public Announcement Logic

Public announcement logic is an extension of static epistemic logic with simple epistemic events, announcements, that transmit true information reliably and publicly (to all agents) (Baltag and Renne 2016). For our purposes, we can think of single-agent PAL as perhaps the simplest logical system that follows the semantic idea presented in Figure 2.1. As an extension of epistemic logic, we have a modal propositional operator K, representing the propositional knowledge of the agent. We further have an update operator $[!\varphi]$ for every φ of the language, s.t. $[!\varphi]\psi$ is read "as a result of the announcement of (or update with) φ, ψ is the case." Diamond duals of the modal operators are defined as usual: $\hat{K}\varphi$ is defined as $\neg K\neg\varphi$ and $\langle !\varphi\rangle\psi$ is defined as $\neg[!\varphi]\neg\psi$.

A formula φ of PAL is evaluated over a Kripke model $M = (W, R, V)$, where W is a non-empty set of possible worlds, R is an epistemic indistinguishability relation that is assumed to be reflexive,[4] and V is a valuation function, mapping every atomic formula to a subset of W. Formulas are evaluated with respect to a pair (M, w) of a Kripke model and a specific point $w \in W$ as standard in modal logic. In particular, the formula $K\varphi$ is true at a possible world w iff all worlds u that are accessible to w via R are worlds in which φ is true. Formally:
$M, w \models K\varphi \;\;\Leftrightarrow\;\; \forall u : wRu, M, u \models \varphi.$

The semantic idea behind the $[!\varphi]$ operator is exactly the one from Figure 2.1: updating the Kripke model with φ results in a model in which all the not φ worlds are eliminated. Since PAL updates are assumed to be veridical, $[!\varphi]\psi$ is taken to be vacuously true in a world in which φ is false, while $\langle !\varphi\rangle\psi$ is taken to be false in such world (this sums up the difference between the box and diamond versions of the update operator). More formally, we have the following semantic clause for the update operator:
$M, w \models [!\varphi]\psi \;\;\Leftrightarrow\;\; M, w \models \varphi \;\textit{implies}\; M_\varphi, w \models \psi.$

The antecedent of the right-hand side of this condition guarantees that when we try to update with φ is a world in which φ is false, $[!\varphi]\psi$ is vacuously true. The consequent states that ψ must be true in the model which results from updating M with φ, which is denoted as M_φ and defined as $(W_\varphi, R_\varphi, V_\varphi)$, where W_φ is just $W \cap \{w \in W \mid M, w \models \varphi\}$ and R_φ and V_φ are the restrictions of R and V with respect to W_φ. With relation to the discussion around Figure 2.1, we can think of $[\cdot]_\varphi$ as a function sending us from Kripke models to Kripke models, s.t. $[M]_\varphi = M_\varphi$. Note that

[4] We are not assuming that R is transitive or Euclidean.

the value of such function is indeed insensitive to the world of evaluation; we don't need an actual world to compute $[M]_\varphi$ from M.[5]

For a simple example, consider Ann, who looks at a tower clock and learns that the time is 12:00 (let that be proposition p). The models in Figure 2.2 depict the PAL update with such p.

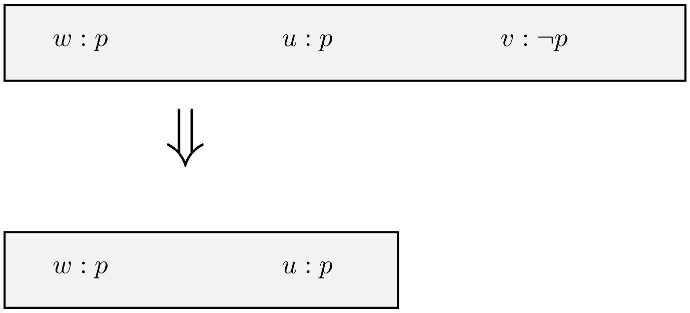

Figure 2.2: A standard PAL update

Figure 2.2 depicts a transition from the prior model M at the top to the posterior model M_p at the bottom. The epistemic R relation is depicted with the grey box. Note that since world v does not satisfy p in the initial model, it is eliminated from the posterior model M_p. In order to compute M_p we don't need to know which world is considered actual. The PAL model transition in Figure 2.2 is just one example of the semantic idea from Figure 2.1.

Reduction axioms for PAL

A popular way to axiomatize PAL is via the following set of *reduction axioms*[6]

1. $[!\varphi]p \leftrightarrow (\varphi \to p)$ atomic reduction
2. $[!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$ negation reduction
3. $[!\varphi](\chi \wedge \psi) \leftrightarrow ([!\varphi]\chi \wedge [!\varphi]\psi)$ conjunction reduction
4. $[!\varphi]K\psi \leftrightarrow (\varphi \to K[!\varphi]\psi)$ knowledge reduction
5. $[!\varphi][!\psi]\chi \leftrightarrow [!\varphi \wedge [!\varphi]\psi]\chi$ announcement reduction

For every formula $[!\varphi]\psi$ of PAL, a repeated application of the reduction axioms results in an equivalent expression ψ^* in which the update operator has been eliminated. In other words, every

[5] In the current presentation of PAL, the world of evaluation does determine whether the partial function $[\cdot]_\varphi$ is defined or not. This is the only sense we can say that the partial function $[\cdot]_\varphi$ is sensitive to the world of evaluation.

[6] Note that for each such axiom, the complexity of β in the sub-expression $[!\alpha]\beta$ is reduced in the right-hand side as compared to the left-hand side (although the overall complexity of the right-hand side expression increases). See Baltag and Renne (2016) for more information.

posterior epistemic state can be syntactically manipulated into an expression about the initial epistemic state. This is similar to the fact that every posterior probability function can be syntactically manipulated into a conditional prior probability function in the Bayesian framework. The common idea is that since the prior epistemic state *determines* every posterior state, every posterior state can be reduced back to the prior state.

2.2.2 Forest models for PAL

While the standard semantics and axiomatization of PAL is quite straightforward, it is not incredibly helpful in highlighting the transparency property of PAL updates. The world elimination semantics do not leave much wiggle room for modifications with the effects of updates, and the reduction axioms do not illuminate any particular principle about the agent's knowledge of the effects of updates, nor do they directly correspond to any semantic frame condition, like many other modal axioms. Still, for any Boolean formula β it is easy to see (either syntactically or semantically) that $K[!\beta]K\beta$ is valid in standard PAL: the PAL agent has the prior knowledge that any update with β results in knowledge that β – there is no uncertainty as to the reliability of (Boolean) updates, they always work.[7]

In order to break the rules of PAL, we will start by presenting an alternative but equivalent semantics for PAL, and a set of axioms that fits nicely with the alternative semantics. The alternative system can be then easily tweaked to construct opaque updates that violate the principles of PAL. The alternative semantics is well-known in the dynamic epistemic logic literature; the novelty of this chapter lies in the way we expand that semantics to analyze opaque updates.

The idea of the alternative PAL semantics is to treat the epistemic result of an update not as a new epistemic model, but as a different part of one big model that contains all its updated models as submodels. Very informally, the idea is to make the meta-theoretic arrow in Figure 2.1 into a regular object-level relation of some Kripke model. On such semantics, the interpretation of the modal expression $[!\varphi]\psi$ is the familiar truth of ψ in all the relevant worlds. Such semantics have been studied extensively in van Benthem et al. (2009) and Wang and Cao (2013).

In order to evaluate $[!\varphi]$ as a regular modal operator, we need to be able to expand a given Kripke model with a $\rightarrow_\varphi$ relation for every formula φ and add new possible worlds for the result of updates. If w is part of the initial epistemic model, and φ is a sentence true in w, then we add the world (w, φ) to the model. World (w, φ), which represents the situation resulting from learning φ in w, has the same atomic valuation as w, and it is connected to w via the $\rightarrow_\varphi$ relation. Such an expanded model is known as the *forest* model of the original model. Here is one way to construct such a forest model from a given epistemic model M (modified from Yap and Hoshi (2009), Aucher and Herzig 2010). The definition is followed by an informal explanation.

[7] The formula $K[!\varphi]K\varphi$ is not valid in general in PAL, however. This is because of Moorean sentences like $p \wedge \neg Kp$ which can change truth value after the announcement. Nevertheless, the PAL agent has no uncertainty as to the reliability of their information sources.

PAL Forest Model: Given a reflexive Kripke model $M = (W, R, V)$ we define its PAL forest $Forest(M) = (F(W), F(R), \to_\varphi F(V))$ s.t.
–$F(W) = \bigcup_n W^n$
–$F(R) = \bigcup_n R^n$
– $\to_\varphi = \bigcup_n \to_\varphi^n$
–$F(V) = \bigcup_n V^n$,
where $M^n = (W^n, R^n, \to_\varphi^n V^n)$ is defined inductively as follows:
– $M^0 = M$, where $\to_\varphi^0 = \emptyset$.
– $M^{n+1} =$

- $W^{n+1} = W^n \cup \{(w, \varphi) \ : \ w \in W^n, \varphi \in \mathcal{L}_{PAL} \ \& \ M^n, w \models \varphi\}$
- $R^{n+1} = R^n \cup \{(w, \varphi), (v, \psi) \ : \ wR^n v \ \& \ \varphi = \psi\}$
- $\to_\varphi^{n+1} = \to_\varphi^n \cup \{w, (w, \varphi) \ : \ w \in W^n\}$
- $V(p)^{n+1} = V(p)^n \cup \{(w, \varphi) \ : \ (w, \varphi) \in W^{n+1} \ \& \ w \in V(p)^n\}$

Here is an informal description of the construction in the definition. We want to create a relation for every possible PAL update φ. $Forest(M)$ is the union of all M^n models, where M^0 represents the original model we start with, M^1 represents the model after one update with φ, M^2 after two updates with φ and so on. Starting at a world w in M^0, if φ is false in that world, then, since PAL updates are veridical, we don't connect any $\to_\varphi$ to that world w. If φ is true at w, we add a new world to the forest model, the world (w, φ), which is part of W^1. That world has the same atomic valuation as w. We then connect w to (w, φ) with the $\to_\varphi$ relation. We extend the epistemic R relation to model M^1 s.t. if w was accessible to u in model M^0, then (w, φ) will be accessible to (u, φ) in model M^1 (assuming that the two worlds exist in M^1). Since we can repeatedly update with φ again and again, the forest construction is infinite.

Figure 2.3 contains a simple example of a forest model.

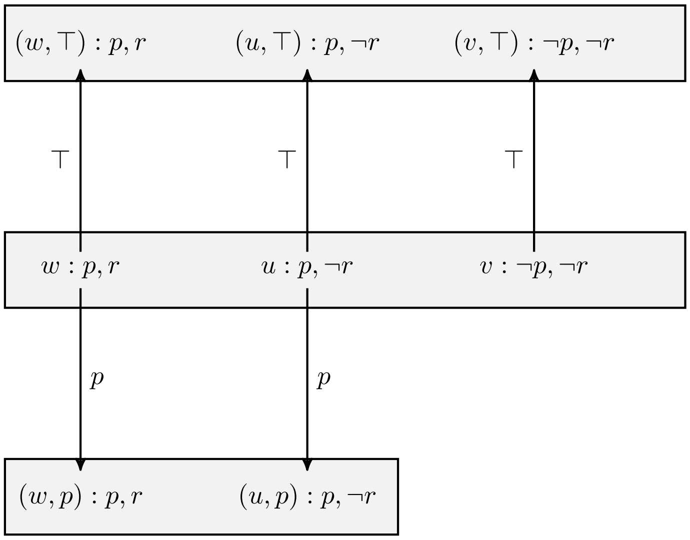

Figure 2.3: A (partial) forest construction

The model in Figure 2.3 (partially) depicts a forest construction, where the middle box is the initial Kripke model from which we extend to a forest construction. The epistemic R relation is an equivalence relation represented by the grey boxes (worlds in the same grey box are connected with the R relation). The black arrows represent the update relations $\to_p$ and $\to_\top$. The $\to_p$ relation depicts the update with the propositional letter p, the result of that update can be seen in the lower submodel. The $\to_\top$ depicts the result of updating with a tautology, which has no effect and results in a copy of the original model, represented in the upper submodel.[8]

The alternative evaluation (denoted $\models_a$) of PAL updates is very simple on forest models: we just follow the $\to_\varphi$. Considering Figure 2.3, we can see that $M, w \models_a [!p](Kp \wedge \neg Kr)$ by following the single p arrow from w and noting that $M, (w,p) \models_a Kp \wedge \neg Kr$; after the update with p, the agent knows p but does not know r.

2.2.3 An alternative axiomatization of PAL

The $\models_a$ evaluation is equivalent to the standard PAL evaluation from Section 2.1 on Forest constructions (Wang and Cao 2013), but, as we will soon see, the alternative semantics on the forest construction can be quite illuminating when it comes to opaque updates. With the alternative semantics, Wang and Cao (2013) have shown that the following set of axioms axiomatizes PAL as well:[9]

6. $(p \to [!\varphi]p) \wedge (\neg p \to [!\varphi]\neg p)$ Atomic Invariance

[8] The model does not depict other updates, like the update with r or $\neg p$, nor does it depicts iterated updates with p or $\top$, although it is clear that such updates have reached a fixed point.

[9] Together with a distribution (K) axiom and a necessitation rule for the operator $[!\varphi]$.

7. $\langle !\varphi\rangle\psi \leftrightarrow \varphi \wedge [!\varphi]\psi$ Partial Function

8. $\langle !\varphi\rangle K\psi \rightarrow K[!\varphi]\psi$ No-Miracles

9. $K[!\varphi]\psi \rightarrow [!\varphi]K\psi$ Perfect-Recall

This way of presenting the axioms of PAL is valuable both because each axiom corresponds to a clear frame condition on forest models and because each axiom (but especially 7.-9.) represents *a non-trivial, debatable, epistemological commitment.*

The atomic invariance axiom states that epistemic updates do not change the non-epistemic facts in the world. Semantically, it corresponds to the conditions that if $x \rightarrow_\varphi y$ then x and y agree on every atomic formula. The axiom defines the events we analyze as *epistemic* events, rather than *ontic* events that change the non-epistemic facts of the world.

The Partial Function axiom essentially states that updates are *deterministic*, that given a particular situation, there is a fact of the matter as to how the update affects the agent (even if the agent does not know that). For a recent discussion about deterministic updates in a Bayesian context, see Pettigrew (2019). Semantically, the axiom corresponds to the fact that each world in the forest model has *at most* one $\rightarrow_\varphi$ coming out of it (for each φ). In my epistemological application of opaque updates (Section 2.3), I assume that updates are deterministic.

The No-Miracles and Perfect-Recall principles

The last two axioms, No-Miracles (NM) and Perfect-Recall (PR), are of the most importance when it comes to understanding the difference between transparent and opaque updates.[10] One way of thinking about NM and PR is as *dynamic introspection principles*, which contrasts them to the well-known, and well debated, static introspection principles like $K\varphi \rightarrow KK\varphi$ and $\neg K\varphi \rightarrow K\neg K\varphi$ (positive and negative introspection, respectively). Static introspection principles involve what we know about our own mental states; syntactically, such principles involve the scoping of the K operator over other instances of the K operator. Analogously, I suggest we call NM and PR dynamic introspection principles because they involve what we know about our own epistemic events. Syntactically, dynamic introspection principles involve the scoping of the K operator over the update operator.[11]

When it comes to ignorance about updates, two independent types of questions arise: future (or forward) directed and past (or backward) directed. The future directed question asks, given a particular event, how is that event going to epistemically affect me: *where am I going from here?*

[10] The No-Miracles principle is not related to the No-Miracles argument from philosophy of science. A better name for NM might be No-Surprises, since, informally, the principle expresses the idea that the agent is never surprised by the result of an update. Here we follow the epistemic logic literature and stick with the name NM. For more on NM and PR in epistemic logic, see van Benthem et al. (2009), van Benthem (2012), Wang and Cao (2013); for the connection with game theory, see van Benthem (2014), van Benthem and Klein (2018).

[11] NM and PR are logically independent from axioms 4. and 5. of epistemic logic (static positive and negative introspection, respectively.) Any combination of dynamic and static introspection is therefore possible.

The NM principle ($\langle !\varphi\rangle K\psi \rightarrow K[!\varphi]\psi$) answers this question in the following way. It states that if the update with φ actually results in the agent being in a position to know ψ ($\langle !\varphi\rangle K\psi$), then the agent has the *prior* knowledge that φ updates result in ψ being the case ($K[!\varphi]\psi$). Taking NM as an axiom amounts to assuming that the agent always has the ability to correctly predict the effect of updates, that the agent is never ignorant as to the result of the update. This is one part of what it means for an update to be transparent to the agent.

Past directed ignorance about updates can be summarized in the question: given my current epistemic position, what update exactly has brought me to this position? *How did I get here?* The PR principle ($K[!\varphi]\psi \rightarrow [!\varphi]K\psi$) offers an answer. PR states that if the agent has the prior knowledge that the update with φ results in a ψ state ($K[!\varphi]\psi$), then as a result of the $[!\varphi]$ update, the agent is in a position to know ψ ($[!\varphi]K\psi$). Consider the negation of PR, stating $K[!\varphi]\psi \wedge \neg[!\varphi]K\psi$. It exemplifies opaqueness towards the update φ. Assume that ψ is the 'mark' of a φ update. The agent knows that ψ is the mark of a φ event, but they don't know ψ after the φ update. This implies that the agent does not know that the update was a φ update. As we will see, however, the failure of PR does *not* imply that the update failed to convey information.

The *opaqueness* terminology for updates is also related to the familiar discussions about opaque contexts from philosophy of language. Here is an unoriginal example of an opaque context, cashed out in terms of a failure of the Perfect-Recall principle. Lois Lane knows that as a result of seeing Superman she is having a special day. But Lane cannot identify Clark Kent as Superman, although they are identical. As a result of seeing Clark Kent, Lois Lane does not know that she is having a special day, even if she knows that as a result of seeing Superman, she is having a special day.

Lois Lane's situation exemplifies a failure of PR. Let $[\psi]$ denote the event of seeing Superman, and let φ be the proposition *Lois Lane is having a special day.* Then both $K[\psi]\varphi$ and $\neg[\psi]K\varphi$ are true, which is a counter example to the PR principle $K[\psi]\varphi \rightarrow [\psi]K\varphi$. $K[\psi]\varphi$ is true because Lois Lane has the *de-dicto* knowledge that seeing Superman is a special event; $\neg[\psi]K\varphi$ is true because Lane is not able to identify Clark Kent as Superman, thus she does not know that a special event just occurred. No forgetfulness is involved on Lane's part, but Perfect-Recall fails. There is a way to describe the epistemic event of seeing Superman s.t. Lane does not know that it is the event that brought her to her posterior epistemic state, namely as the event of seeing Clark Kent.

Both NM and PR can be challenged on the grounds of the cognitive limitations of actual, non-idealized, agents. Rejecting PR as a way of modeling an agent's memory loss has been discussed within Bayesian epistemology, but I am not aware of analogous discussions about NM.[12] A different

[12] See, e.g., Arenzious (2003) and Halpern (2004) for discussions about PR. Traces of the underlying commitments behind NM and PR can be detected in the Bayesian conditionalization scheme $P_H(E) = P(H|E)$, where P_E stands for the posterior probability function after learning (with certainty) E. If we break down conditionalization into two directions for a specific numerical constant c, say $c = 1$, we get the commitments:
(i) $P_E(H) = 1 \;\Rightarrow\; P(H|E) = 1$
(ii) $P(H|E) = 1 \;\Rightarrow\; P_E(H) = 1$
(i) and (ii) roughly correspond to instances of NM and PR, respectively. (i) states that if the agent actually becomes certain in H after the update with E (in the sense of assigning probability 1), then the agent has the *prior*

question is the compatibility of NM and PR with various theories in epistemology, regardless of cognitive limitations; the next section of this work is devoted to such issues.

NM and PR, as modal axioms, correspond to frame conditions on forest models. NM states that if wRt and $t \to_\varphi v$, and there is a world u s.t. $w \to_\varphi u$, then uRv. PR states that if $w \to_\varphi u$ and uRv, then there is a world t s.t. wRt and $t \to_\varphi v$ (Wang and Cao 2013). The two semantic conditions can be elegantly summarized in the following figure:

$$\begin{tikzcd}
u \arrow[r, "R"] & v \\
w \arrow[u, "\varphi"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
u \arrow[r, "R"] & v \\
w \arrow[u, "\varphi"] \arrow[r, "R"] & t \arrow[u, "\varphi"]
\end{tikzcd}
\qquad
\begin{tikzcd}
u & v \\
w \arrow[u, "\varphi"] \arrow[r, "R"] & t \arrow[u, "\varphi"]
\end{tikzcd}$$

PR ⇒ ⇐ NM

Figure 2.4: NM and PR as semantic conditions

Informally, NM and PR together imply that forest models are commutative in the following sense: every world you can get to by first going via the R relation and then by the φ relation, you also get to by first going via the φ relation and then the R relation, and vice versa. Consult the forest model in Figure 2.3 for an example.

2.2.4 Adding opaque updates to PAL

We are now in a position to see how to construct updates that break the NM and PR properties. The idea I am going to present is to *compose* new update relations from existing PAL update relations. Consider again Figure 2.3. If we had a tool to pick subsets of the union of the $\top$ and φ relations, we could easily create new relations (updates) that don't respect the commutative structure implied by NM and PR. Fortunately, there exists a modal tool that allows us to reason about composition of relations, known as propositional dynamic logic, or PDL. Combining PDL with dynamic epistemic

certainty that H is the case, given E. (ii) states that if the agent starts with the prior certainty in H given E, then that certainty is not lost once E is actually learned. However, unlike NM and PR in PAL, which connect two dynamic expressions, conditionalization connects a dynamic expression (the posterior state) with a static attitude (prior conditional probability).

logic is a rich field of study.[13] In what follows, I present a simple way to combine the alternative forest semantics of PAL with a fragment of PDL to get a logic that is sufficiently flexible to model opaque updates.

Unlike PAL, the logic of opaque updates denotes updates with expressions π that themselves are not always wffs of epistemic logic. Such $[\pi]$ operators represent updates that might be basic PAL updates or some composition of PAL updates. The language of the logic of opaque updates is defined inductively as follows:

$$\varphi := \;\; p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid K\varphi \mid [\pi]\varphi$$

$$\pi := \;!\psi \mid ?\psi \mid \pi_1; \pi_2 \mid \pi_1 \cup \pi_2$$

where the formula ψ is of the language of epistemic logic (i.e. does not contain the update operator).[14] In PDL, π is called a program. $!\psi$ is considered in the logic of opaque updates as an atomic program (from which we compose more complicated programs). Every PAL update is an atomic program in the logic of opaque updates. For example, $!p$ is the program whose execution amounts to a PAL update of the formula p. $?\psi$ is a test program, which checks if ψ is true and aborts otherwise. $\pi_1; \pi_2$ is a composite program that sequentially executes program π_1 and then π_2. $\pi_1 \cup \pi_2$ is a composite program that non-deterministically executes π_1 or π_2. The syntax of PDL is useful for expressing familiar algorithmic expressions: the expression "if α, do π_1, otherwise do π_2" is written as $(?\alpha; \pi_1) \cup (?\neg\alpha; \pi_2)$. In the next section we will see epistemological examples of such expressions. The expression $[\pi]\varphi$ reads "after every execution of π, φ is the case." The expression $\langle\pi\rangle\varphi$ reads "there is an execution of π after which φ is the case."

In what follows, we evaluate formulas of the logic of opaque updates over forest models. Given some epistemic Kripke model M, we evaluate φ on $Forest(M)$. The semantic clause for the K operator is the same as in epistemic logic. The semantic clause of the expression $[\pi]\varphi$ is straightforward:
$M, w \models [\pi]\varphi \;\; \Leftrightarrow \forall u : wR_\pi u, M, u \models \varphi$,
where the relation R_π is composed inductively in the following manner:

- $wR_{!\psi}u$ iff $w \rightarrow_\psi u$,
- $wR_{?\psi}w$ iff $Forest(M), w \models \psi$,
- $wR_{\pi_1;\pi_2}u$ iff there is a v s.t. $wR_{\pi_1}v$ and $vR_{\pi_2}u$,
- $wR_{\pi_1 \cup \pi_2}u$ iff either $wR_{\pi_1}u$ or $wR_{\pi_2}u$.

[13] See Troquard and Balbiani (2019) for a survey of PDL. See, e.g., van Ditmarsch (2000), van Benthem et al. (2006), Wang et al. (2009), Girard et al. (2012) for works connecting PDL and DEL.

[14] The restriction on ψ simplifies the technical discussion with no effect on the philosophical part. We exclude the Kleene star operator of full PDL from this system. For the study of PAL with a Kleene star for updates, see Miller and Moss (2005).

For a few examples, consider $Forest(M)$ from Figure 2.3 again. The relation $R_{!p\cup!\top}$ is just the union of the $\rightarrow_p$ and $\rightarrow_\top$ relations in that model, and the relation $R_{?(p\wedge r);!p}$ only consists of the pair of worlds $(w,(w,p))$, since, out of w,u, and v, only world w satisfies $(p \wedge r)$.

2.2.5 Reduction axioms for opaque updates

The logic of opaque updates is completely axiomatized by adding to the PAL axioms the following PDL axioms:

10. $[?\psi]\varphi \leftrightarrow (\psi \rightarrow \varphi)$

11. $[\pi_1;\pi_2]\varphi \leftrightarrow [\pi_1][\pi_2]\varphi$

12. $[\pi_1 \cup \pi_2]\varphi \leftrightarrow [\pi_1]\varphi \wedge [\pi_2]\varphi$

Note that axioms 10.-12. are reduction axioms, meaning that every formula ψ of the logic of opaque updates can be reduced to a formula ψ^* in the language of PAL (i.e., a formula that has no updates with $?, ;$ or $\cup$). For example, the formula

$$[(?r;!p) \cup (?\neg r;!\top)]Kp$$

can be equivalently written as (using 10.- 12.)

$$(r \rightarrow [!p]Kp) \wedge (\neg r \rightarrow [!\top]Kp),$$

which is a PAL formula, and any PAL formula, as we already discussed, can itself be reduced to a formula without any update operators. In fact, we can think of the equivalences in 10. to 12. as the definitions for the syntactic abbreviations $?, ;$ and $\cup$. This shows that the forest constructions (and the alternative semantics that follows it) which we have used are not strictly necessary: we can think of any program π in the logic of opaque updates as an abbreviation of a PAL expression, which can be evaluated with the standard PAL semantics.

Like in PAL (and Bayesian updating for that matter), in the logic of opaque updates, every formula expressing a posterior epistemic state can be rewritten as an expression without an update, describing the agent's prior state. The difference is that in the logic of opaque updates the truth value of the reduced sentence will not only depend on the epistemic situation, but also on non-epistemic factors (e.g., the truth of r in the above example). The prior situation still determines everything, but it must include more than just the prior *epistemic* situation.

Obviously, NM and PR will not be valid principles in a logic designed to model opaque updates. We can formulate the latter principles in our richer language as

13. $\langle\pi\rangle K\varphi \rightarrow K[\pi]\varphi$ No-Miracles

14. $K[\pi]\varphi \rightarrow [\pi]K\varphi$ Perfect-Recall

The semantic conditions of NM and PR, depicted in Figure 2.4, can also be generalized by replacing the φ relations with π relations. The next section presents concrete counterexamples to formulas 13. and 14., motivated by recent debates in contemporary epistemology.

2.3 Opaque updates in contemporary epistemology

This Section discusses a few applications of opaque updates in debates in epistemology. The applications center around issues concerning externalist conceptions of knowledge, broadly understood. The below Subsections are not meant to settle the philosophical discussions one way or the other, rather to show that opaque updates, and the NM and PR principles, are useful and relevant conceptual tools in such discussions.

2.3.1 Basic knowledge theories

Following Cohen (2002), we use the term *basic knowledge theories* to refer to epistemological theories that do not reject the possibility of *basic knowledge.* Basic knowledge, in this context, is the knowledge that is obtained from a reliable source without the agent having the prior knowledge that the source is indeed reliable. The term *basic knowledge theories* is meant to be an umbrella term that applies to many approaches to knowledge, most of which are considered externalist or naturalist in one way or another. Examples for such theories include reliabilism, safety theories, sensitivity theories, anti-luck theories, and causal theories of knowledge.[15] Common to these theories is the idea that the agent does not need to know that a source of information is reliable in order to actually gain knowledge from that source.

Let us complicate our simple example from Figure 2.2 to the following toy example.

> **Clock Tower:** Ann is visiting a foreign village. She looks at a clock tower that points to 12:00. The clock tower has, in fact, a perfectly reliable clock mechanism, although Ann does not know that (neither before looking at the clock nor after).

Basic knowledge theories will gladly accept that it is possible for Ann to come to know that the time is 12:00 as a result of looking at the clock (that would be basic knowledge).

Theories that reject basic knowledge argue that since Ann does not know that the clock mechanism is reliable, she does not come to know the time. Both sides of this debate agree that if that clock mechanism is in fact unreliable, then Ann does not come to know the time by looking at the clock.

[15] For a survey of these theories, see Ichikawa and Steup (2018). See also Lyons (2016) for a related discussion about *modest foundationalism.*

We can take a dynamic perspective on this debate. Both sides agree on an initial epistemic situation in which Ann does not know that the time is 12:00 (denote that proposition p) and does not know that the clock mechanism is reliable (denote that with r). The basic knowledge theorist describes the event of looking at the clock as the event s.t. if the clock mechanism is reliable (r) then Ann comes to know p, while if the clock is unreliable then she comes to know nothing new. The theorist that rejects basic knowledge describes the event of looking at the clock as the event in which no new knowledge about the time is gained, no matter what (because, as both sides assumed, Ann does not know that the clock mechanism is reliable).

It is worth highlighting here the way many externalists have come to understand the notion of *evidence*, and the way it functions in examples like the one above. Contrast two scenarios: in the first the clock is broken (and hence unreliable); in the second the clock is reliable. Intuitively, when Ann looks at the clock that points to 12:00, she receives the same evidence in both scenarios. Externalists tend to reject this intuition. According to many contemporary externalists, evidence is factive: evidence is always true and false evidence is not real evidence.[16] If evidence is factive, then Ann does not receive the same evidence in both scenarios: if the clock is in fact broken, Ann does not have evidence that the time is 12:00; if the clock is reliable, she does. Externalists defend the claim that evidence is factive by offering an 'external' analysis of evidence, under which the status of evidence is affected by factors that go beyond the agent's intrinsic mental state (like whether the source of the information is reliable or not). Externalists can accept that in both scenarios Ann comes to believe that the time is 12:00, even if she has different evidence in the two scenarios. This picture, of course, complicates the relation between one's belief, evidence and knowledge, and remains a topic of debate within epistemology.[17] The agent's doxastic state will not be modeled in this section.

In the clock tower example, we assume that Ann does not know if the clock is reliable or not. Hence, according to the externalist, she cannot predict what she will learn (or what evidence she will receive) as a result of looking at the clock. I claim that we can model the externalist understanding of the event of Ann looking at the clock as the program $(?r; !p) \cup (?\neg r; !\top)$. We read this program as the following event: if r is the case (the clock mechanism is reliable), transition into a situation in which Ann knows p (the time is 12:00); if r is not the case, transition into a situation in which Ann has not learned anything new. Figure 2.5 depicts this type of update with a forest model.

[16] See Williamson (2000), Littlejohn (2013), Bird (2018), Neta (2018), Fratantonio and McGlynn (2018), and Salow (2017) for a factive take on evidence. See Goldman (2009), Rizzieri (2011) and Comesaña and Kantin (2010) for objections. My models do not strictly speaking assume that evidence is factive, because the models do not explicitly model evidence. But I do think that the models help frame this debate.

[17] This problem has been labeled the *new evil demon problem*. For an overview, see Littlejohn (2009) and Beddor and Goldman (2015).

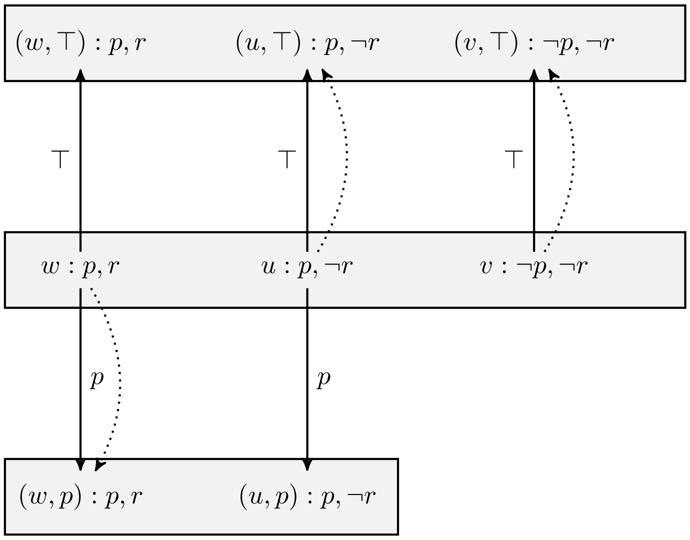

Figure 2.5: A basic knowledge conception of learning p without knowing that the source is reliable (the dotted relation).

The middle box in Figure 2.5 represents Ann's initial situation: she does not know p (that the time is 12:00) nor does she know r (that the clock mechanism is reliable). The dotted relation depicts the program (or event) $(?r;!p) \cup (?\neg r;!\top)$, which is the event of learning p if r is the case and learning nothing if not r is the case. Observe the depiction of that event, as the dotted relation, in Figure 2.5: in the world in which r is true (world w, the clock mechanism is indeed reliable), the dotted relation sends us to a submodel in which p is known (at the bottom). In the worlds in which the clock mechanism is unreliable (worlds u and v) the dotted relation sends us to a submodel in which nothing has changed (the top model). The event of looking at the clock has different epistemic consequences, depending on which world is actual.

A few observations about the dotted relation, the epistemic event $(?r;!p) \cup (?\neg r;!\top)$: first, note that it does not fit the framework depicted in Figure 2.1, it is an update which is sensitive to the world of evaluation (the external environment); it is not just a function from one model to the other. Second, note that this event is deterministic: at each particular environment it has a deterministic result.[18] Graphically, the relation is a partial function (a function in fact). Third, note that the dotted relation violates the semantic condition which corresponds to the NM axiom: from w we can get to (w,p) via the dotted relation, from w we can get to u via the R relation (which, recall, is depicted as the grey boxes), and from u we can get to $(u,\top)$ via the dotted relation. But we cannot get from world (w,p) to world $(u,\top)$ via the R relation, as the semantic NM condition would require.

Thus, the event $(?r;!p) \cup (?\neg r;!\top)$ depicted in Figure 2.5 offers a counter example to the NM

[18] Thus, the opaqueness of that event can only be attributed to the agent's ignorance, not to any ontic chances.

axiom, stating that $\langle\pi\rangle K\varphi \to K[\pi]\varphi$. On the one hand, it is the case that $M, w \models \langle(?r; !p) \cup (?\neg r; !\top)\rangle Kp$: in world w, as a result of looking at the clock, Ann comes to know that the time is 12:00 (according to the basic knowledge theorist). On the other hand it is also the case that $M, w \models \neg K[(?r; !p) \cup (?\neg r; !\top)]p$: prior to looking at the clock, Ann does not know that looking at a clock pointing to 12:00 implies that the time is in fact 12:00, since, recall, Ann does not know that the clock mechanism is reliable. For all Ann knows, the actual world is v, in which the time is not 12:00 and the clock mechanism is unreliable. In such a world, looking at a clock that points to 12:00 does not imply that the time is 12:00. Thus, the basic knowledge theorist that understand the event of looking at the clock as the dotted relation rejects the NM principle.[19]

The event $(?r; !p) \cup (?\neg r; !\top)$ is opaque: prior to its execution, Ann does not know what its result will be. Ann does not know if she is in the good case or the bad case, so the effect of looking at the clock is not transparent to her. The event fails the NM principle because Ann does not know where she is going, epistemically speaking.

The objector to basic knowledge uses the contra-positive form of NM to reject the possibility of an event like $(?r; !p) \cup (?\neg r; !\top)$. According to the objector, since Ann does not know that the event π results in a p state (you don't know that looking at a clock that says the time is 12:00 implies that the time is actually 12:00), then the event π does not result in knowledge that p (you don't really know that the time is 12:00 after looking at the clock). This is the implication $\neg K[\pi]p \to \neg\langle\pi\rangle Kp$, which is the contra-positive form of NM, $\langle\pi\rangle Kp \to K[\pi]p$. We can understand the objector to basic knowledge as endorsing NM, and with it, a transparent conception of updates (if you don't know that the source is reliable, you can't get knowledge out of it).

Upshot: since the rejection of NM allows agents to gain knowledge from sources they don't know to be knowledge conducive, the truth of NM plays a key role in debates about basic knowledge.[20] PAL with opaque updates allows us to model both sides of this debate.

2.3.2 The bootstrapping problem

The bootstrapping problem is a famous objection that is usually raised against basic knowledge theories.[21] According to objectors, basic knowledge theorists allow agents to learn about the reliability of their sources of information for free, or too easily, after the update, without sufficient evidence

[19] As I mentioned in the introduction, externalism is often assumed to be inconsistent with the KK principle. See Okasha (2013), Bird and Pettigrew (2019) for recent discussions and Williamson (2000) for the locus classicus. The picture painted here complicates the discussion. I have argued that it is natural to understand externalists as rejecting NM first and foremost. Since KK and NM are logically independent, there is nothing inconsistent about rejecting the one while endorsing the other. I believe that it is worthwhile re-evaluating familiar arguments against introspective knowledge given the distinction between static and dynamic introspection I draw here. This is done in Chapter 5 of this dissertation.

[20] The philosophical connections between No-Miracles, basic knowledge, and skepticism are further developed in Chapter 3 or Cohen (2020).

[21] See Vogel's attack on reliabilism (2000, 2008) and Cohen (2002). Some take the problem to be more widespread (Weisberg 2010, van Cleve 2003). Analogous problems exist for justification and belief (White 2006, Pryor 2013, Weatherson 2007).

for doing so. This way of learning amounts to illicit bootstrapping, or easy knowledge, according to the objector. The bootstrapping problem comes in many variations and has many interesting proposed solutions.[22] The discussion here is not meant to offer a decisive solution to the problem, but to show that opaque updates, and in particular the PR principle, are quite relevant to at least one formulation of the problem.

We can present a toy version of the problem with our clock example modeled in Figure 2.5.[23] The objector to basic knowledge theories notes that initially (in the middle submodel of Figure 2.5) Ann knows that as a result of looking at the clock, if she comes to know the time, then the clock mechanism must be reliable. In other symbols, the formula $K[\pi](Kp \to r)$ is true at world w, where $\pi = (?r; !p) \cup (?\neg r; !\top)$, the event of looking at the clock according to the basic knowledge theorist. There is no world in the model in Figure 2.5 accessible via the R_π relation in which $Kp \wedge \neg r$ is true, so we get $M, w \models K[\pi](Kp \to r)$. Now, the objector argues informally as follows: we agree that Ann has the prior knowledge that as a result of looking at the clock, if she knows the time, then the clock mechanism must be reliable. Further, according to the basic knowledge theorists, Ann can come to know the time after actually looking at the clock. Thus, what stops Ann from putting the two pieces of information together and concluding, after looking at the clock, that the clock mechanism is in fact reliable (illicitly bootstrapping her way to the knowledge that the clock is reliable, so to say)? The problem is that in order to learn that the clock mechanism is reliable, it seems, one needs to do more than merely glance at it. Since the conclusion that Ann comes to know that the clock is reliable just by looking at the clock face is absurd, something must be wrong with the informal argument. The objector suggests rejecting the basic knowledge theorist's assumption that Ann learns the time by looking at the clock.

Here is one way to make the above informal argument more formal: we assume $[\pi]Kp$ (Ann knows the time as a result of looking at the clock) and $K[\pi](Kp \to r)$ (as in the last paragraph). Using the PR principle, $K[\pi]\varphi \to [\pi]K\varphi$, we can conclude $[\pi]K(Kp \to r)$, stating that as a result of looking at the clock, Ann knows that if she knows the time, the clock mechanism must be reliable. Given the KK principle and the closure principle of knowledge, from $[\pi]Kp$ and $[\pi]K(Kp \to r)$ we can deduce $[\pi]Kr$: as a result of looking at the clock, Ann knows that the clock is reliable. To summarize:

15. $[\pi]Kp$ basic knowledge assumption

16. $K[\pi](Kp \to r)$ assumption about Ann's background knowledge

17. $[\pi]K(Kp \to r)$ by PR from 16.

18. $[\pi]KKp$ instance of KK on 15.

[22] See, e.g., the survey in Weisberg (2012).

[23] This example of bootstrapping does not involve any inductive elements, unlike Vogel's original version of the problem (2000, 2008). However, as Titelbaum (2010) has argued, induction is not essential to the bootstrapping problem.

19. $[\pi]Kr$ by the closure of knowledge from 17. and 18.

So, my formulation of the bootstrapping argument requires assuming an instance of the closure of knowledge, the KK principle, and the PR principle.[24] Note that the model in Figure 2.5 makes sentences 15. and 16. true (at world w) and assumes both the KK principle (the transitivity of the R relation) and the closure principle of knowledge, but $[\pi]Kr$ is false: after looking at the clock (world (w,p)) Ann does not know r. The bootstrapping argument is blocked in Figure 2.5 because the update π, apart from violating NM, further violates PR. From world w one can go first to world (w,p) via the π relation (dotted), then to world (u,p) via the R relation, but there is no way to get from world w to world (u,p) first via the R relation and then via the dotted π relation. Syntactically, note that $K[\pi](Kp \to r) \wedge \neg[\pi]K(Kp \to r)$ (which is true in M, w) is a counterexample to the PR principle.

The model in Figure 2.5 violates the PR principle not because Ann forgets anything after looking at the clock. The failure of PR occurs because Ann, even after the epistemic event occurred, does not know what event has brought her to her current epistemic situation. Consult Figure 2.5 again, and assume that w is the actual world. We can see that both $K[\pi](Kp \to r)$ and $[\pi]Kp$ are true. However, after the event of looking at the clock (world (w,p)), Ann considers it possible that the actual world is (u,p), meaning she considers it possible that the initial actual world is u and that she learned that p at that world. This type of ignorance can be cashed out as a difference between *de-re* and *de-dicto* knowledge. Ann can have the de-dicto knowledge that as a result of the event *looking at a clock tower*, if she knows the time, then the clock mechanism must be reliable. At the same time, she can be ignorant of the fact that *that* event is the event that resulted in her current knowledge (this is *de-re* ignorance). The ignorance expressed in the failure of PR arises from the possibility that different distinct reliable sources could have given the agent the knowledge they have. For a concrete, but somewhat outlandish example, take world u to be a world in which the clock mechanism is unreliable but in which a benevolent God transplants the correct time in Ann's mind. Since Ann cannot rule out this non-actual epistemic event (maybe she just read Descartes' *Meditations*), she cannot rule out the possibility that an epistemic event different than the actual one produced her knowledge. The epistemic event that Ann experienced is opaque in the sense that she cannot identify it as the event from which she obtained knowledge about the time. This is why world (u,p) cannot be ruled out after the update.[25]

Existing discussions about the bootstrapping problem seem to be unaware of the fact that PR plays an essential role in bootstrapping reasoning. We see that externalists can dissolve the bootstrapping problem by noting that an update like the clock tower update is opaque, and therefore

[24] The formulation also assumes that $[\pi]$ is a normal modal operator. Recall that we indeed treat π as a normal modal operator in the logic of opaque updates. I don't see how this assumption would be challenged on epistemological grounds.

[25] Further note that if world (u,p) would be eliminated from the model in Figure 2.5, then PR would be true in w, while NM would still be false. One way of formally interpreting the bootstrapping objection is as the insight that rejecting NM while endorsing PR is a strange epistemological combination.

incompatible with PR. In general, PR will fail in situations in which agents don't know how they know φ (or what is the epistemic event that resulted in knowing φ), even if they know φ. Again, basic knowledge theories are more open to the possibility of such situations, so such theories could be understood as rejecting both PR and NM. The simple PAL forest model of Figure 2.5 can represent this complicated epistemic scenario.

2.4 Comparison to other frameworks and further work

Opaque updates, as an epistemic-logical phenomenon, are not exclusive to the particular logical system I have presented here. Sections 2.2 and 2.3 presented and used a simple PDL-PAL hybrid system that has the advantage of being relatively close to the (now very) familiar PAL, while flexible enough to easily model the failure of both NM and PR. In this section, I discuss other logical frameworks that highlight interesting features of opaque updates:

1. Multi-pointed event models: One standard way to generalize PAL is to model updates via *event models* that can represent a wide range of communication events. In DEL with event models, an updated epistemic model is the product $M \times A$ of an initial epistemic model M and an event model A. An event model is a Kripke model with a finite set of events E, an accessibility relation between events, and a precondition function for each event $e \in E$, which intuitively specifies which formula is announced for each e. An event modality $[A, e]\varphi$ is added to the language, stating that φ holds as a result of executing event e. See Baltag and Renne (2016) for a proper overview on event models.

Consider the event model A with two events, e_1 and e_2, such that the precondition of e_1 is p and the precondition of e_2 is $\top$. In the context of the clock tower example, we can think of $[A, e_1]$ as the event of looking at a reliable clock (resulting in knowledge of p) and of $[A, e_2]$ as the event of looking at an unreliable clock (resulting in no new knowledge). $[A, e_1]$ and $[A, e_2]$ are not opaque updates, intuitively because they specify which event is actual (e_1 or e_2). We can, however, easily build opaque updates out of them by using the *multi-pointed* event operator $[A, E]$ (see Sietsma and van Eijck (2012), Baltag and Renne (2016)).[26] The *multi-pointed* event operator $[A, E]\varphi$ abbreviates the conjunction $\bigwedge_{e \in E} [A, e]\varphi$. Likewise, the diamond operator $\langle A, E\rangle$ stands for the disjunction $\bigvee_{e \in E} \langle A, e\rangle\varphi$.

The operator $\langle A, E\rangle$ is non-deterministic and opaque, since the agent cannot know which event e_i it executes. Taking the middle model of Figure 2.5 as our initial epistemic model, we have the following No-Miracles failure: $M, w \models \langle A, E\rangle Kp$ and $M, w \models \neg K[A, E]p$. Note, however, that there is no direct equivalence between this approach to opaque updates and the one I presented in Sections 2.2-2.3. Recall that conceptually, there is a difference between whether an update is deterministic and whether the agent can predict its behavior. The latter is an epistemic issue; the former is ontic. The program π from Section 2.3 is deterministic, even though the agent cannot predict its

[26] I thank an anonymous reviewer of the journal of philosophical logic for suggesting this connection.

behavior. The update $\langle A, E\rangle$ on the other hand is truly non-deterministic (note that we have $M, w \models \langle A, E\rangle \neg Kp \wedge \langle A, E\rangle Kp$), which explains why the agent cannot predict its behavior. This difference also implies that we cannot directly interpret the formula $K[A, E]\varphi \rightarrow [A, E]K\varphi$ as capturing the epistemic intuition behind Perfect-Recall (in the clock example, the sentence $K[A, E](Kp \rightarrow r) \rightarrow [A, E]K(Kp \rightarrow r)$ is true as its antecedent is false). A richer study of opaqueness using the framework of event models is of course welcomed, and I leave it for future investigation.

2. The dynamics of knowing a value: There is a growing literature on dynamic-epistemic logics for wh-knowledge, like knowing *what* is the value of a variable.[27] Consider the event of announcing the truth value of p. This event is opaque in that it violates No-Miracles: if p is true and the agent does not know it, the agent cannot predict that they will know p after the event of announcing the truth value of p. The agent can predict, however, that after the event they will know the value of p, so they are not ignorant about the fact the event is successful in conveying the value of p (this is unlike the examples from Section 2.3). Epistemic logics that go beyond knowing-that can provide further stimulating perspectives on opaqueness. See Cohen et al. (2021) for this direction.

3. Plausibility models for doxastic epistemic logics: Existing dynamic doxastic epistemic logics offer an impressive model for formalizing an externalist notion of knowledge as belief that is stable under revision with true information. The framework of Baltag and Smets (2008) in particular, succeeds in combining different notions of belief, knowledge, update, and belief revision in a unified manner. The *radical upgrade* operator $\Uparrow \varphi$ of that system represents the agent's belief revision with the (possibly false) φ, and is modeled with a doxastic plausibility ordering. The system includes two notions of knowledge: an S4 type modality K representing defeasible knowledge, and a stronger S5 modality $\Box$ representing irrecoverable knowledge. Defeasible knowledge is equivalent to belief that is stable under revision with true information.

It is quite interesting to study opaqueness relative to different modalities and updates in this system. For example, the NM principle $\langle \Uparrow \varphi\rangle K\psi \rightarrow K[\Uparrow \varphi]\psi$ is not valid in the system: an agent will always come to know p after a radical upgrade with p, assuming that p is true, i.e. $p \rightarrow \langle \Uparrow p\rangle Kp$. But if the agent initially does not believe that p, then they will not know that revising p implies that p is actually the case: $\neg Bp \rightarrow \neg K[\Uparrow p]p$. This failure of NM nicely captures the externalist elements of the defeasible knowledge operator K: the epistemic result of an update depends on factors the agent is initially ignorant about.

Plausibility models also provide an excellent tool to further study the effects of false evidence from an externalist perspective. I have used the program $\pi = (?r; !p) \cup (?\neg r; !\top)$ to model the clock tower example. This modeling is incomplete, since it ignores the (quite plausible) assumption that regardless of the reliability of the clock (the truth of r), the agent comes to believe the time is 12:00 (i.e. p) as a result of looking at the clock (even if they don't know p). We can model the clock tower example with plausibility models, and consider a program like $\pi' = ((?r; !p) \cup (?\neg r; !\top); \Uparrow p)$. This

[27] See Wang (2018) for a broad overview, and van-Eijck et al. (2017), Baltag (2016) for dynamic epistemic logics of knowing a value.

program will have different effects on the agent's irrecoverable knowledge, but it results in belief in p come what may.[28]

4. Topological models of epistemic logic: The opacity of updates relate to interesting topological properties in topological semantics of epistemic logic, when updates are understood as functions defined over topological spaces.

To see the connection, let us transform the Kripke structures we have been using into topological models (see Section 1.4.3 for an overview on topological models). Given an **S4** Kripke frame (W, R), an *upset* A, $A \subseteq W$ is such that if $x \in A$ and xRy then $y \in A$. Defining a topology τ on W by having every upset be open (i.e. in τ) results in an Alexandroff topology in which for every $w, w \in W$, the set $R(w) = \{u \in W : wRu\}$ is a least neighborhood of w. In general, there is a 1-1 correspondence between Alexandroff spaces and **S4** frames. (see van Benthem and Bezhanishvili 2007: 30-31).

Since the Kripke model in Figure 2.5 is both finite and **S5** (with respect to the epistemic R relation) its corresponding topological model is just the model with a topology such that each grey box represents a smallest non-empty open set. So, for example, the smallest open set containing w is the set $\{w, u, v\}$.

Given this topological interpretation of the epistemic model in Figure 2.5, we can further interpret the programs $!p, !\top$ and π from Figure 2.5 as the *partial functions* $f_{!p}, f_{?p}, f_{\pi}$ defined over our topological space. This is because the relations $\rightarrow_{!p}, \rightarrow_{!\top}$ and $\rightarrow_{\pi}$ are partial functions. In general, note that any atomic program (either $!\varphi$ or $?\varphi$, where $\varphi \in EL$) is indeed a partial function by the construction of the forest model: if $f_{!p}(w) = x$ and $f_{!p}(w) = y$ then $x = (w, p) = y$ by the forest construction, as needed. Another way to see that is by noting that the axiom Partial Function (discussed in Section 2.2.3) is sound on Forest constructions for every $\rightarrow_{!\varphi}$ relation (Wang and Cao 2013), so every relation $\rightarrow_{!\varphi}$ is indeed a partial function on Forest models.

Note however, that it is not true in general that all update programs that can be composed in the logic of opaque updates are partial functions. Consider the composed program $\pi' = !p \cup !\top$ relative to the model in Figure 2.5. π' is not a function: we have both $w \rightarrow_{\pi'} (w, \top)$ and $w \rightarrow_{\pi'} (w, p)$ for such a relation. Some composed programs are partial functions. The program $\pi = (?r; !p) \cup (?\neg r; !\top)$, depicted with the dotted lines in Figure 2.5, is an example of a partial function.

Given an **S4** frame (W, R), a function f is an *order preserving map* if xRy implies $f(x)Rf(y)$. Given a topological space, a function f is *continuous* at point x if, for every open neighborhood V of $f(x)$, there exists an open neighbourhood U of x such that $f(U) \subseteq V$.

Recall that as a semantic condition on Forest models the No-Miracles principles states that if wRt and $w \rightarrow_{\varphi} u$ and $t \rightarrow_{\varphi} v$, then uRv (Figure 2.4 summarizes this fact). Since we can treat the relation $\rightarrow_{\varphi}$ as the partial function f_{φ}, we can restate No-Miracles as stating the following: if

[28] As I briefly mentioned in the introduction, I chose to model opaque updates with the simple PDL-PAL hybrid system of Section 2.2 because (unlike plausibility models) it remains neutral with respect to the truth of the KK principle and the exact relations between knowledge and belief. Both of these issues are controversial in contemporary philosophical discussions of externalism.

xRy and f_φ is defined on points x and y, then $f_\varphi(x)Rf_\varphi(y)$. In other words, No-Miracles states that updates, when understood as partial functions, must be map preserving on every point they are defined over.

The 1-1 correspondence between Alexandroff spaces and **S4** frames implies that there is a 1-1 correspondence between order preserving maps (on **S4** frames) and continuous maps (on the corresponding Alexandroff spaces) (see van Benthem and Bezhanishvili 2007: 31). This, together with the above remarks, immediately implies that we can treat No-Miracles as stating that on topological models, the update partial function f must be continuous wherever it is defined. Since standard PAL updates validate No-Miracles, it follows that standard PAL updates are continuous wherever they are defined.

With the model in Figure 2.5, we can also show that opaque updates, like the dotted relation in the figure, are *not* always continuous. The dotted relation, representing the program $\pi = (?r; !p) \cup (?\neg r; !\top)$, is a partial function (a function in fact, on the model in Figure 2.5), so we can think of it in terms of f_π. Note that f_π is defined over point w of the model in Figure 2.5. Continuity requires that for every open neighborhood V of $f_\pi(w)$, there exists an open neighbourhood U of w such that $f_\pi(U) \subseteq V$. Take V to be the open set $\{(w,p),(u,p)\}$ in the Alexandroff topology corresponding to the model in Figure 2.5. Note that as $f_\pi(w) = (w,p)$, V is in fact an open neighborhood of $f_\pi(w)$ (i.e., $f_\pi(w) \in V$). Since the only open neighbourhood U of w has to be the open set $\{w,u,v\}$, we need to check that $f_\pi(U) \subseteq V$, i.e. $f_\pi(U) \subseteq \{(w,p),(u,p)\}$. But as $f_\pi(U) = \{f_\pi(w), f_\pi(u), f_\pi(v)\} = \{(w,p),(u,\top),(v,\top)\}$, we get that $f_\pi(U) \not\subseteq V$, contrary to what continuity requires. The opaque update π of Figure 2.5 is not continuous under a topological interpretation, in contrast to standard, transparent, PAL updates, which are.

It is worthwhile to further study the relationship between opaque and transparent updates and topological semantics in a more systematic manner, and I defer this to future work. See Kremer and Mints (2005), and Bjorndahl (2018) for relevant existing works that connect continuity on topological semantics of modal logic with dynamic aspects of modal logic.

2.5 Concluding remarks

Taking a dynamic perspective, we can think of epistemic internalism as the claim that posterior epistemic states *supervene on the agent's prior epistemic state*. Indeed, in some formal frameworks (like Bayesian update), sentences about the posterior epistemic state can be *reduced* to sentences about the agent's prior epistemic state. Externalists, on the other hand, hold that the effects of epistemic events supervene both on the epistemic state of the agent *and* on environmental conditions external to the agent. Opaque updates allow us to model the externalist idea that the result of an epistemic event may depend on non-epistemic features of the environment. An opaque update U can result in knowledge that p in a world where the non-epistemic fact r (which is logically independent of p) is true and can also result in no new substantial knowledge in a world in which r is false. As we have seen, the semantics of such opaque updates directly correspond to the failure of the (syntactic) NM and PR principles. It is thus valuable to understand externalist theories as rejecting NM and PR.

From our current perspective, it is fair to say that Hintikka's (1962) seminal work in epistemic logic has *reshaped* the way many epistemologists think about introspection. Static positive and negative introspection are now taken for granted as tools in the epistemologist's toolbox, whether people accept them or not. My hope is that, likewise, recent developments in dynamic epistemic logic will offer epistemologists the conceptual tools to think about *dynamic* forms of introspection in a precise and simple manner.

Chapter 3

The problem of perception and the no-miracles principle

Chapter abstract: The problem of perception is the problem of explaining how perceptual knowledge is possible. The skeptic has a simple solution: it is not possible. I analyze the weaknesses of one type of skeptical reasoning by making explicit a dynamic (or diachronic) epistemic principle from dynamic epistemic logic that is implicitly used in debating the problem, with the aim of offering a novel diagnosis to this skeptical argument. I argue (i) that prominent *modest foundationalist* responses to perceptual skepticism can be understood as rejecting the dynamic assumption made by the skeptic, (ii) that there are independent reasons to doubt the truth of such a principle in the context of skeptical reasoning, and (iii) that making the dynamic principle explicit allows for a better understanding of at least one objection to modest foundationalism.

The problem of perception, in its epistemic guise, is the problem of explaining how perceptual knowledge is possible. The skeptic has a simple solution: it is not possible. In this chapter, I want to offer a novel analysis of the skeptical line of reasoning. I argue that a *diachronic introspection principle* plays, implicitly, a key role in the skeptical argument, and that understanding its nature is beneficial for our understanding of the overall skeptical problem of perception.

Diachronic (or dynamic) introspection principles are, roughly, principles from *dynamic epistemic logic* that describe how epistemic events affect the subject's epistemic state, given the subject's initial epistemic state. Although such principles are familiar within various traditions in formal epistemology, they have not received enough attention in epistemology as a whole. This is unfortunate since—at the very least—such principles may allow us to elucidate and understand disagreements in epistemology, in particular, disagreements regarding the problem of perception, as I shall argue here.

I start, in section 3.1, by presenting the diachronic introspection principle that will be at the focus of this chapter, the *No-Miracles* principle, and its context within game theory and formal epistemology. In section 3.2, I reconstruct the problem of perception as assuming the No-Miracles principle and argue that the skeptic should be understood as implicitly endorsing it. I further argue that prominent responses to the skeptical problem can be characterized as rejecting the No-Miracles principle. In section 3.3, I consider independent reasons for rejecting the No-Miracles principle in the face of the skeptical challenge. Instead of trying to answer the skeptic, these considerations aim to offer a new diagnosis of the intuitive appeal of skepticism about perceptual knowledge. Section 3.4 reconsiders an objection made by White (2006) in light of my reconstruction of the problem of perception. I argue that by making the role of the No-Miracles principle explicit, the objection can be better understood. Section 3.5 concludes.

3.1 The No-Miracles principle

Diachronic introspection principles can be represented as schemas which different epistemic events and states can be plugged into. The epistemic state that will be at the focus of this work is propositional knowledge, but in principle diachronic introspection principles can be applied to any number of epistemic states. Let $\Box$ stand for some epistemic or doxastic state that takes propositions, e.g. believing that..., being certain that..., assigning probability c to that..., knowing that..., etc. Let E be a propositional operator describing some epistemic event. $E\psi$ stands for the expression 'as a result of the epistemic event E, ψ is the case.' Thus, sentences of the form $E\psi$ are able to describe a certain *dependency* between an event E and a state of affairs that follows it, expressed in ψ. This chapter is not meant to study the essence of this dependency, but it does assume its existence. After all, talk of *perceptual knowledge* assumes a certain dependency: a connection between a perceptual event and a knowledge state. Note that the same point holds for *testimonial knowledge* and *inferential knowledge*: these types of knowledge assume that there is some dependency between an event (testimony, inference), and a resulting knowledge state. E can in principle stand for *any* event that has the power to change the subject's epistemic state: learning that φ is the case, suffering from hypoxia, receiving the reliable testimony that φ, getting the evidence that φ, etc. Our focus in this chapter is on epistemic events that relate to perceptual knowledge; in particular, we focus on the epistemic event of *having the experience as of* p, which figures prominently in discussions regarding the epistemological aspects of perception. Once we go on to discuss the problem of perception exclusively (from section 3.2 of this chapter onward) we can understand the term *epistemic event* as a placeholder for the expression *having the experience as of* p, for some particular p and some particular agent who experiences p. For now, in the rest of this section, we will consider a few examples of epistemic events of different type, to highlight some features of the No-Miracles principle.[1]

[1] The answer to the question what counts as an epistemic event for an agent will depend in general on context. For example, in the context of a game of perfect information, the event of one player playing their turn will count as

To give an example as to how this event notation is used, let E_p denote the event of having the experience as of p. Then the expression E_pKp states that as a result of having an experience as of p, the subject comes to know p. Likewise, $\neg E_pKp$ states that the subject does not come to know p as a result of the experience (maybe because the experience was part of a hallucination). And the expression E_pp states that having the experience as of p results in a p state, which is just to say that the experience is veridical; similarly $\neg E_pp$ would express that the experience is not veridical.

The diachronic introspection principle we will focus on in this work is the No-Miracles principle. Its general schematic form is $E\Box\varphi \rightarrow \Box E\varphi$, but our main focus here is on knowledge, so we can take the principle to be the following:

No-Miracles (NM): $EK\varphi \rightarrow KE\varphi$.

The principle states that if as a result of event E the subject is in a position to know φ, then the subject is antecedently in a position to know that event E results in φ being the case. The name *No-Miracles* comes from the epistemic logic and game theory literature. A better name would be *no-surprises*, since the principle intuitively express the idea that the agent is never surprised by the outcome of an epistemic event, as I will explain below.[2] Here I will follow the existing epistemic logic literature and stick with the name No-Miracles.[3] As with other epistemic principles (e.g. the closure of knowledge principle, the KK principle), we can read expressions of the form $K\varphi$ as stating that the agent is *in a position to know* φ, instead of *the agent knows* φ; this allows us to abstract away from some of the cognitive limitations of actual agents.[4] For the purpose of this chapter, it suffices to understand the implication in the principle as the material implication.[5]

For a simple example of applying NM, consider the following state of a tic-tac-toe game:

O	O	
	X	
	X	

an epistemic event for the players of the game, as we assume (in such context) that all players are perfectly informed about every turn in the game. The same event might not be an epistemic event for a different agent who is not part of the game.

[2] The connection I draw here between the No-Miracles principle and surprise is not meant to be an exhaustive account of surprise in epistemic logic. For different types of analyses of surprise in epistemic logic, see Lorini and Castelfranchi (2007), Demey (2015) and Demey and Vignero (forthcoming).

[3] The principle should not be associated with the no miracles argument from philosophy of science, nor with any other philosophical discussion about miracles for that matter. For a review of the principle in epistemic logic, see van Benthem et al. (2009), van Benthem (2012), and Wang and Cao (2013). For its role in game theory, see van Benthem and Klein (2019). It is worth noting that NM is the converse of a more familiar diachronic principle, known as *Perfect-Recall*, which can be stated as $KE\varphi \rightarrow EK\varphi$. Versions of No-Miracles and Perfect-Recall can be used as axioms of the dynamic epistemic logic *Public announcement logic* (PAL), as I have shown in Chapter 2. This alternative axiomatization is studied and compared to the standard reduction axioms of PAL in Wang and Cao (2013).

[4] By using the expression *being in a position to know* we just guard against counterexamples to the NM principle that are based on the cognitive limitations of actual human subjects. If someone is distracted or not paying attention, they might not know φ, even if they are in a position to know φ. The way I use this term follows Williamson's usage, according to which "If one is in a position to know p, and one has done what one is in a position to do to decide whether p is true, then one does know p" (2000: p 95).

[5] Since much of this chapter criticizes an unqualified NM principle, it is most charitable to consider the weakest interpretation of it.

Suppose that X's turn comes next. Given the current state of the game, if X plays bottom left, then O is going to be in a position to know that they can immediately win, by playing top right. It follows that O is already in a position to know that if X plays bottom left, then O can immediately win. In other words, the fact that O can win if X plays bottom left should not come as a surprise to O. It is entirely predictable given the *current* state of the game.

This kind of informal reasoning pattern exemplifies the NM principle. To see why, let E stand for the event *X plays bottom left*, and let $K_O\varphi$ stand for *O is in a position to know* φ. Given the current state of the game, the formula

$$EK_O(\text{O can immediately win})$$

is true. It just states that as a result of X playing bottom left, it is true that O is in a position to know that they can immediately win the game. From this, we have argued that the following must be true as well:

$$K_OE \text{ (O can immediately win)},$$

i.e., O is already in a position to know that as a result of X playing bottom left, they (player O) can immediately win the game. Note that the conditional

$$EK_O \text{ (O can immediately win)} \rightarrow K_OE \text{ (O can immediately win)}$$

is an instance of the NM principle, expressing the idea that once X actually plays bottom left, O does not experience any surprise (i.e., no epistemic miracle occurs), since O could already tell what would be the outcome of such action. In general, the NM principle is assumed in game theoretic scenarios of perfect information in which players can only get new information by observing publicly available actions (van Benthem and Klein 2019).

The NM principle is also fundamental (albeit, implicitly so) to Bayesian Epistemology. The most basic diachronic postulate of Bayesian Epistemology, the norm of conditionalization, states that $P(H|E) = P_E(H)$, i.e. the prior probability of H conditional on E should equal the posterior probability of H, where the posterior probability is the new probability function of the agent after learning E. The norm of conditionalization implies instances of the NM principle. Consider the special case of assigning probability 1. The norm of conditionalization then implies:

$$P_E(H) = 1 \Rightarrow P(H|E) = 1$$

i.e. if once one learns E, one becomes certain of H, then one is certain (in the sense of assigning probability 1) in H conditional on the truth of E. The idea here is that the Bayesian agent is never truly *surprised* when they become certain of a proposition H after learning E, as they can always *antecedently* correctly predict that conditional on E, they are certain of H. Letting $\Box^1\varphi$ stand for the propositional operator *assigning probability 1 to* φ, we can rewrite $P_E(H) = 1 \Rightarrow P(H|E) = 1$ as

$$E\Box^1(H) \to \Box^1 E(H)$$

where $E(H)$ is understood to mean that given E, or as a result of learning E, H, and $E\Box^1$ as representing the propositional operator *assigning posterior probability 1.* The conditional $E\Box^1(H) \to \Box^1 E(H)$ says that if after the event of learning E you are certain of H then you are a-priori certain of E given H; it is an instance of the NM schema (where knowledge is replaced with probabilistic certainty). In general, we can never think of the Bayesian agent as being unable to predict the result of a given learning event, since the result, i.e. the posterior probabilities, is given by the prior conditional probabilities. Ideal Bayesian agents obey the NM principle, at least on straightforward understanding of the theory.[6,7]

Viewed syntactically, the NM principle, $EK\varphi \to KE\varphi$, permits us to move from a narrow scope reading of the knowledge operator over φ (in the antecedent) to a wide scope reading over the E operator (in the consequent). There is a long tradition of understanding the difference between wide and narrow scope readings of an intensional attitude like knowledge as correlating to the difference between *de-re* and *de-dicto* knowledge (Nelson 2019). One way of reading NM is as saying that if an event E results in the agent knowing that φ, then it must be the case that the agent has the *de-dicto* knowledge that E results in φ. Viewed this way, counterexamples to the principle can be easily generated.

Here is a counterexample to NM which is generated by the *de-dicto* ignorance the agent has about an event (under a particular description). Suppose that Alice accidentally dropped a candle on her couch on the 11th of November 2011, at 11:11. As a result, she came to realize that the couch is on fire.[8] So it is true that a result of the thing that happened to Alice on 11.11.11 at 11:11, Alice knows that the couch is on fire, because, as it happens, the thing that happened to Alice on 11.11.11 at 11:11 is the event of accidentally dropping the candle on the couch. But it does not mean that Alice *knows* that the thing that happened to Alice on 11.11.11 at 11:11 resulted in the couch being on fire, because Alice might not know that she dropped the candle on 11.11.11 at 11:11 (say she is unaware of the time). If E stands for *the thing that happened to Alice on 11.11.11 at 11:11*, and φ stands for *the couch is on fire*, it might very well be that $EK_{Alice}\varphi$ while $\neg K_{Alice}E\varphi$, which is a counterexample to NM. Since Alice does not know that the description *the event of dropping the candle* and the description *the thing that happened to Alice on the 11.11.11 at 11:11* refer to the same event, she does not have the *de-dicto* knowledge that the thing that happened to Alice on

[6] Things are more complicated. Agents might be 'surprised' when we consider (i) Jeffrey conditionalization instead of standard conditionalization and (ii) when updating on event with probability 0 occurs. See Joyce (2003) for some overview.

[7] Failures of the NM principle can lead to the failure of the probabilistic *diachronic reflection principle* (Briggs 2009, Arntzenius 2003). Informally, you might be uncertain whether your future self is more informed than your current self, even if your future self is indeed more informed. In such cases both NM and diachronic reflection will fail. Counterexamples to diachronic reflection are more familiar, however, when a different diachronic introspection principle fails, namely *Perfect-Recall*.

[8] We do not assume here that the dependency between the event and Alice's epistemic state can be completely analyzed in causal terms.

11.11.11 at 11:11 resulted in the couch being on fire (i.e. $\neg K_{Alice} E\varphi$). A full commitment to NM amounts to rejecting the possibility of reasoning about such non-transparent events.

I take the NM principle to be a *diachronic* or *dynamic* principle, but it is important to note that considerations about actual time can be mostly abstracted away. I treat NM as a diachronic principle because it is stated in a language that extends epistemic logic (the K operator) with an explicit reference to epistemic events, via the E operator. In this sense, a principle like NM should be contrasted with principles that I consider synchronic (or static), like the KK principle, $K\varphi \rightarrow KK\varphi$, which only uses the K operator. The important thing about having both reference to knowledge (the K operator) and to epistemic events (the E operator) is the ability to describe a *dependency* between epistemic events and certain states of affairs. As I already mentioned, *perceptual knowledge* is a prime example of such dependency: knowledge that *results* from, or *depends on* a perceptual event. Although it is very natural to think of such dependency in a temporal fashion, it is not necessary for our understanding of NM.[9]

A final remark: this chapter is not meant to offer any kind of analysis of the notion of dependency which is mentioned in the NM principle. First of all, the nature of this dependency will change according to the type of event and epistemic state in question. Even when we focus on perceptual knowledge, i.e., the knowledge state that depends on an event of perceptual experience, the nature of this particular dependency will vary widely across different theories of perceptual knowledge. I do not assume that the dependency between an epistemic event and the resulting epistemic state is causal dependency, and nothing in this chapter hinges on such assumption. The point of this chapter is not to endorse one such theory of the dependency or the other. There is no question, however, that some dependency is assumed in a concept like perceptual knowledge. I take it as primitive and focus on the reasoning patterns which are associated with it. Moreover, my aim is not to defend or reject the NM principle. We have already seen how to generate counterexamples to it. The point is to recognize and understand the reasoning pattern that the NM principle captures. Unfortunately, we can still implicitly use a principle that we explicitly reject. The formal mechanisms I use will hopefully help in seeing how to uncover a principle like NM.

3.2 The problem of perception

How does all of this relate to the problem of perception? In his formulation of the problem, which I will follow here, Lyons (2017) flags two premises crucial for the skeptical argument.[10] These are what he calls the **Reasons claim** and the **Metaevidential principle**:

- **Reasons claim:** We have no good reason for thinking perceptual appearances are veridical.

[9] For a concrete example, consider the tic-tac-toe game from before. Even if in fact both players will *never* touch the board again, it is still the case that EK_O(O can immediately win), where E is the event *X plays bottom left.* Thus, E should not be understood as denoting a particular moment in time. E is not a temporal modality per-se.

[10] He also considers a more basic principle, the **Indirectness principle**, stating that nothing is directly presented to the mind. See the discussion in (Lyons 2017: sec 1.)

– **Metaevidential principle:** Without a good reason for thinking perceptual appearances are veridical, we are not justified in our perceptual beliefs (Lyons 2017: sec. 1).

The skeptic endorses these two premises and concludes that we are not justified in our perceptual beliefs. The truth of the **Reasons claim** is supported, according to Lyons, by the fact that we cannot rule out the possibility of illusion. The **Metaevidential principle** is supported by the observation that "if our access [to the world] is mediated by potentially non-veridical appearances, then we should only trust the appearances we have reason to think veridical" (ibid).

Although I would like to follow the general argument structure presented by Lyons, two small changes to the presentation are in order. First, my focus will be on local instances of the skeptical reasoning, not (directly) about the universal generalization. Second, I wish to focus on perceptual *knowledge*, not about justified beliefs. Taking these two changes into account, it is natural to present the skeptical premises with the following two sentences, given a particular perceptual appearance:

(1) The subject does not know that this particular perceptual appearance is veridical.

(2) Without knowing that this perceptual appearance is veridical, the subject does not have perceptual knowledge in this instance.

The skeptic I will focus on concludes that the subject does not have perceptual knowledge in the particular case for which premises (1) and (2) hold. Global skepticism about perceptual knowledge then amounts to applying these two premises to each instance of potential perceptual knowledge.

We are now in a position to see the connection between this type of skepticism and the NM principle. To make things concrete, let p stand for the proposition that there is a pelican in the sky. Consider E to be the event of *having the experience as of* p, and abbreviate it as $[exp(p)]$.[11] $[exp(p)]$ refers to a particular event in which the subject has the experience as of p, say when the subject stands on Rodeo beach during sunset and seems to see a pelican in the sky. We use the phrase *experience as of* p to make clear that such experiential events (or appearances) are *not* assumed to be veridical. The skeptical argument can then be formulated as follows. The first premise of the skeptical argument is that the subject does not know that an experience as of p results in p being the case (because the subject cannot rule out the possibility of non-veridical experience). In symbols:

(I) $\neg K[exp(p)]p$

The purpose of (I) is to abbreviate premise (1) from earlier, a version of what Lyons calls the **Reasons claim**. The conclusion of the skeptical argument is simply $\neg[exp(p)]Kp$. In English: it is not the case that the experience as of p results in knowledge that p. To get from to (I) to the skeptical conclusion one just needs to assume the implication

[11] More accurately, we should distinguish between the event $exp(p)$ and the propositional operator $[exp(p)]$. $exp(p)$ just abbreviates the event of having the experience as of p; the propositional operator $[exp(p)]$ allows us to formulate what happens as a result of the update.

$\neg K[exp(p)]p \to \neg[exp(p)]Kp$.
Note that this implication is the contrapositive of the **NM** principle. Hence, the connection between (I) and the skeptical conclusion is an instance of **NM**, where $E = [exp(p)]$:

(II) $[exp(p)]Kp \to K[exp(p)]p$.

(II) says that if the perceptual experience as of p results in knowledge that p, then it must be the case that it is known that the experience as of p results in a p state, i.e. the experience is veridical. So, (I) and (II) give us the conclusion that $\neg[exp(p)]Kp$ via modus tollens. Further note that (II) abbreviates in symbols premise (2) from earlier. On my reconstruction of the skeptical argument, the **Metaevidential principle** is just an instance of the No-Miracles principle. The skeptical position regarding perceptual knowledge makes an implicit use of this diachronic principle. Further note that, once NM is accepted, this skeptical argument is surprisingly strong: apart from NM, it just rests on one more premise, namely (I). In particular, it does not assume the closure of knowledge. To conclude, the skeptical argument has the following formal structure:

(I) $\neg K[exp(p)]p$

(II) $[exp(p)]Kp \to K[exp(p)]p$

C: $\neg[exp(p)]Kp$

where C is obtained by modus tollens via (I) and (II). Blocking the skeptical argument amounts to a rejection of either (I) or (II). Rejecting (I), the **Reasons claim**, will not be the focus of this chapter.[12] Instead, I will focus on premise (II), as an instance of **NM**. The question then is: what is wrong with the rejection of NM, the conjunction $\neg K[exp(p)]p \wedge [exp(p)]Kp$, stating that the subject does not know that the p experience is veridical, but as a result of the experience as of p they come to know p?[13]

First, I would like to point out that some prominent responses to the skeptical argument should be understood as rejecting the NM principle. More accurately, some *modest foundationalist* responses to skepticism can be characterized as rejecting the validity of the NM principle. Modest foundationalism is the view that some justified beliefs are basic, in the sense that they are not based on any other beliefs. The reliability of these basic beliefs might be based on something other than beliefs, but the epistemic subject does not need to have justified beliefs about the reliability of these basic beliefs. (cf. Hasan and Fumerton 2016, Lyons 2017). Although the discussion about modest foundationalism usually pertains to justification (in the context of the regress problem of epistemic justification),

[12] Some coherentist and conservative responses to perceptual skepticism can be characterized as rejecting the **Reasons claim**. See Lyons (2017) for this path.

[13] There is nothing implausible about the conjunction $\neg K[exp(p)]p \wedge [exp(p)]Kp$ from the perspective of epistemic logic. If we let $[exp(p)]$ stand for the *radical upgrade* operator $[\Uparrow p]$ from dynamic epistemic logic (Baltag and Smets 2008, van Benthem 2012), and if we interpret K as the S4 *defeasible knowledge* operator of Baltag and Smets (2008), then such a conjunction is indeed satisfiable. See van Benthem (2011), Baltag and Smets (2008), Baltag and Renne (2016), for more about such formal systems.

it is also applicable for knowledge. Modest foundationalists about knowledge claim that some of the subject's knowledge is basic, in the sense that the subject is not expected to be in a position to offer any kind of justification for it (cf. Cohen 2003). I argue that a commitment of modest foundationalist views is a rejection of NM.

Consider externalist theories of knowledge and justification as an example of theories that allow for basic knowledge. Such theories will not accept the truth of $[exp(p)]Kp \rightarrow K[exp(p)]p$. In the externally good case (where the environmental conditions are right, e.g., I am not a brain in a vat) it might very well be the case that $[exp(p)]Kp$. However, since it can also be the case that I cannot know the claim that I am in the good case, $\neg K[exp(p)]p$ might be the case as well. In general, any generic account of knowledge as an external state will concede that it is possible to know p (because one is in the good case) without knowing that one is in the good case. Therefore, under an externalist conception of knowledge, there are all the reasons to think that $[exp(p)]Kp \wedge \neg K[exp(p)]p$ is satisfiable. For a concrete example, consider a generic form of reliabilism. According to reliabilism, given that the subject's vision is reliable, the perceptual experience as of p can result in knowledge that p (i.e. $[exp(p)]Kp$), even if the subject is not in a position to know that the experience as of p is formed in a reliable fashion ($\neg K[exp(p)]p$) (see, e.g. Goldman and Beddor 2016). Reliabilists should reject NM.

For a different modest foundationalist view, consider Jim Pryor's *Dogmatism*. Dogmatism is an example of an internalist, modest foundationalist view of justification (Pryor 2000). According to Dogmatism, perceptual experiences are *prima facie* justified, even if the subject cannot justify the claim that perceptual justification is reliable. Dogmatism is concerned with a response to a skeptical argument about perceptual justification. We can reconsider a variant of the formal reconstruction of the skeptical argument from before, stated in terms of perceptual justification.

(I) $\neg J[exp(p)]p$

(II) $[exp(p)]Jp \rightarrow J[exp(p)]p$

C: $\neg[exp(p)]Jp$

In this argument, the K operator is replaced with the J operator. In this version of the argument, premise (II) is an instance of the NM principle for justification. NM, in this instance, states that in order to gain justification from a perceptual event ($[exp(p)]Jp$) one must be antecedently justified in believing that the experience is veridical ($J[exp(p)]p$). Dogmatists will reject the NM principle for justification. According to Dogmatism, it is possible to gain justification from a perceptual experience (i.e. $[exp(p)]Jp$) even if one is not justified in believing that the experience is reliable, let alone veridical (i.e. $\neg[exp(p)]Jp$). Considering the case of Moore having a perceptual experience as of his own hand, Pryor writes:

> My view is that when Moore's experiences represent there to be hands, that *by itself* makes him prima facie justified in believing there are hands. [...] There are things

> Moore could learn that would undermine this justification. But it's not a condition for having it that he *first* have justification to believe those undermining hypotheses are false (Pryor 2004: 356, original emphases)

Pryor makes it very clear that Moore's perceptual experience can make him justified even if Moore does not have the *antecedent* ability ("it's not a condition ... that he *first* have justification...") to rule out the undermining hypotheses that imply that the experience is non-veridical. Stated in my suggested formalism, Pryor is accepting the possibility that $[exp(hand)]J(hand)$, the experience *makes* Moore justified, even if $\neg J[exp(hand)](hand)$, Moore is not able to first justifiably rule out the non-veridical alternatives. Pryor is rejecting the diachronic NM principle. Note the way that Pryor uses diachronic language (...*first* have justification) in his articulation of Dogmatism.

The third example I wish to consider is Disjunctivism. Here I follow Lyons (2017) in treating disjunctivisim broadly as a modest foundationalist view, although the connection is not as straightforward.[14] Disjunctivism is a prominent view that addresses the problem of perception, and it can be worthwhile to see how the analysis presented here interacts with it. Disjunctivists reject the assumption that a veridical experience of p is of the same kind of as a non-veridical experience as of p (McDowell 1982, 1994, Pritchard 2012). Some further argue that there is no kind of mental state common to veridical and non-veridical experiences (see Byrne and Logue 2008).

Recall that I presented the skeptical argument with the following two premises:

(I) $\neg K[exp(p)]p$

(II) $[exp(p)]Kp \rightarrow K[exp(p)]p$

I have argued that modest foundationalism tends to reject (II). Disjunctivists have *prima facie* reasons to reject premise (I). If by *the experience as of* p we refer to the veridical type of experience, then surely we know, a-priori, that veridical p experience is in fact veridical. Letting $[ver(p)]$ and $[\neg ver(p)]$ stand for the events of veridically and non-veridically experiencing p, clearly $K[ver(p)]p$ is the case. This can be used to reject premise (I). But it is not clear that this observation addresses the core argument of the skeptic (Wright 2002). Even if there is no common element between the events $[ver(p)]$ and $[\neg ver(p)]$, one can still reason about the 'disjunctive' event of either having a veridical experience that p or of having a non-veridical experience as of p. Call the latter disjunctive event $[exp(p)^*]$. For this disjunctive perceptual event, it is still the case that it is unknown to be veridical , i.e. $\neg K[exp(p)^*]p$. So although premise (I) of the skeptic requires some change, it is still compatible with disjunctivism.

However, under the disjunctive conception of perceptual events there seems to be no reason to accept (II), i.e. the NM principle. To see why, note that the implication $[exp(p)^*]Kp \rightarrow K[exp(p)^*]p$, when $[exp(p)^*]$ is understood as a disjunctive event, can refer to one disjunct in the antecedent and another in the consequent. If this is the case, it is unsurprising that the implication can be false. In

[14] See Lyons (2017: sec 3.4.2)

the good case, the disjunctive event $[exp(p)^*]$ can refer to the veridical event $[ver(p)]$. The veridical event is such that it can generate perceptual knowledge. In such a situation, $[exp(p)^*]Kp$ is the case. But the $[exp(p)^*]$ in the consequence can refer to the non-veridical event. In such a case, clearly $\neg K[exp(p)^*]p$.

To conclude this point, neither under the veridical conception of experience of the disjunctivist nor under the disjunctive conception of experience of the disjunctivist is there a reason to accept the two premises of the skeptic. Under the veridical conception, premise (I) is false. Under the disjunctive conception, premise (II) is false.

3.3 The plausibility of No-Miracles

I have argued that prominent modest foundationalist responses to the problem of perception can be characterized as rejecting the NM principle. It is also fair to consider on what general grounds the NM principle can be challenged, in the context of the problem of perception. This section is devoted to such considerations. I argue that NM assumes a form of *epistemic transparency*, and that it is reasonable to question this assumption.

Under the reconstruction I presented, the skeptical argument can be understood as assuming the NM principle. But the skeptic does not have to accept the implication $EK\varphi \rightarrow KE\varphi$ for *any* kind of epistemic event E. As I have argued earlier, it is easy to generate counterexamples to $EK\varphi \rightarrow KE\varphi$ when we can choose any kind of description for E, while taking into account that the agent might be ignorant about the connection between the description and the event. In the context of the problem of perception we care about events of the form *the experience as of* p. We can also consider a demonstrative reference to such events, as in *that* experience as of p (given some fixed particular experience). It suffices for the perceptual skeptic to defend the NM principle when it is restricted only to this type of events. In return, when we wish to challenge this skeptical assumption, we should make the same restriction.

Although I do not intend to offer a knockdown argument against NM, I believe that once we take the skeptic as endorsing it, we have a novel perspective from which we can reevaluate the skeptical reasoning. I take the NM principle to be a diachronic *introspection* principle, because it describes what an agent antecedently knows about an event, given that the event has a certain epistemic effect on the agent. But the event in question is an *epistemic event*, and the epistemic state in question is the agent's own. NM states that assuming that a certain epistemic event has the ability to produce knowledge (i.e. E results in $K\varphi$), the agent is in a position to know something about that event (i.e. the agent knows that E results in φ). Thus, the principle assumes a certain introspective ability of the agent.

But if our representational abilities regarding our own epistemic events are sometimes too coarse-grained to capture important distinctions between events, then we should not expect to have the kind of introspective ability that NM assumes that we have. Suppose that there are two epistemic events of the type of *experience as of* p, E and E', s.t. E is veridical and knowledge conducive (according to a particular epistemological theory) while E' is not veridical (i.e. E implies p, while E' does not). Now further suppose that the agent's initial epistemic state is not able to distinguish between the two given a fixed context:[15] for any property P, for all the agent knows E has P if and only if for all the agent knows E' has P.[16] Since E' is not veridical, it is impossible to know that it is. Thus, $\neg KE'p$. Since the agent cannot epistemically tell E' from E, it must be that $\neg KEp$. The No-Miracles principle now implies that $\neg EKp$ (via modus tollens). This contradicts our assumption that event E can generate knowledge. In other words, the following three assumptions are inconsistent:

(a) There are two epistemic events, E and E', where E can generate knowledge that p while E' is non-veridical (and so cannot generate knowledge).

(b) The agent's prior epistemic state treats E and E' exactly the same.

(c) The NM principle.

Given that E' is in fact non-veridical, for all the agent knows, E' is not veridical. Given (b), for all the agent knows E is non-veridical, which implies that the agent does not know that E is veridical. The NM principle, (c), then implies that it is not the case that event E can generate knowledge, contradicting (a).

The rejection of (a) amounts to radical skepticism about perceptual knowledge, the position that denies that there is any perceptual event E that can generate perceptual knowledge that p. The skeptic I am envisioning here is the one that rejects (a) *on the grounds* of (b) and (c). This skeptic manages to set an extremely high bar for knowledge by endorsing (b) and (c), which imply the negation of (a). In order to have knowledge, this skeptic argues, one must be in a position to tell the E event from the E' event. This kind of reasoning embodies the NM principle. As a response to the skeptic, we note that by endorsing (c), the skeptic *assumes* an extremely high bar for knowledge. This is because, one can argue, what the inconsistency of (a) (b) and (c) shows is that assumption (c), the NM principle, is *extremely* strong, given the prior judgment that (a) and (b) are rather weak. The skeptical argument gains its strength by appearing to move from rather weak, innocuous assumptions to the very strong skeptical conclusion. But by understanding the nature of (c), the NM principle, we can doubt the judgment that such a principle is in fact so innocuous.

Note that non-skeptical externalist or disjunctivist positions will be able to tell a story as to why E and E', although indistinguishable as (b) requires, have different epistemic effects as described in

[15] Fixing a context is important. Let E' be the experience of a hand in the epistemologically bad case, and E the experience of hand in the good case. Of course I can distinguish between E and E' in a context in which they are described the way I just described them: one is the good case, the other is the bad case. But there are other contexts in which I cannot distinguish between the two.

[16] For suppose that there is a property P s.t. the agent knows that E has P while they do not know that E' has P. Then the agent's epistemic state is (or evidence) makes a distinction between the two, contrary to what we assumed.

(a). Reliablism, for instance, can take E to be an event whose source is a reliable process and E' an event whose source is unreliable while at the same time assume that E and E' are indistinguishable from the point of view of the agent. But the exact story behind the acceptance of (a) and (b) is not the heart of the matter here. The acceptance of (a) and (b) can be motivated purely by epistemic modesty considerations. It amounts to the view that sometimes epistemic events are more fine-grained than our coarse grained representation of them. This discrepancy makes reasoning about epistemic events not as transparent as we might assume, but it does not imply skepticism. Endorsing (c), or NM, either amounts to the conclusion that we can tell the two distinct epistemic events E and E' apart (rejecting (b)) or to the conclusion that we don't have perceptual knowledge in such case (rejecting (a)). The skeptic chooses to reject (a) in every case.

It is worth mentioning that a certain type of internalism can endorse NM while avoiding the skeptical conclusion by rejecting (b). Such internalists reject Lyon's **Reasons claim**, arguing that we do have prior reasons to believe that events like event E and, in general, our perceptual faculties, are veridical (Lyons 2017). I have argued that the skeptic endorses NM; it does not follow that those who endorse NM are skeptics.

The NM reasoning pattern is very intuitive, useful and natural in many epistemic contexts. But it would be a mistake to assume without argument that since the NM scheme comes naturally in tic-tac-toe and other simple scenarios it must be applicable in *every* epistemic scenario. If this line of reasoning is correct, then we might have a novel diagnosis of the skeptical argument. The skeptical argument implicitly assumes the NM principle, a diachronic principle which is handy in understanding the epistemic structure of many non-skeptical scenarios. That is not sufficient, however, to conclude that NM is applicable in cases where the skeptical scenario is salient.

3.4 Rethinking the Bayesian objection

Characterizing the modest foundationalist response to the skeptic as rejecting the NM principle also helps in clarifying an existing objection. Roger White (2006) argues that Bayesian epistemology is incompatible with Dogmatism, and thus with modest foundationalism. Following Weisberg's (2015) presentation of the problem, let E stand for the experience of a hand, and H stand for actually having a hand. Hence, $P(E \wedge \neg H)$ denotes the probability of having a non-veridical experience of a hand, and $P(\neg(E \wedge \neg H))$ denotes the probability of the negation of the non-veridical possibility. It follows from the probability calculus that
$P(H|E) \leq P(\neg(E \wedge \neg H))$,
i.e., the conditional probability of H given E is never strictly greater than the probability of not being in the non-veridical possibility. Using the Norm of Conditionalization, we can conclude that the posterior confidence level of having a hand after having the experience of a hand cannot be greater than the prior (or initial) confidence level in the possibility that one is not deceived. From

this, it can be plausibly argued that one cannot really gain justification from perceptual experience unless one is already justified in one's perceptual capacities.[17]

This Bayesian line of reasoning is quite *predictable* under the analysis presented here. Recall that the Bayesian Norm of Conditionalization implies instances of NM. In Bayesian epistemology facts about the reception of new evidence are completely grounded on (indeed, reducible to) facts about the initial epistemic state. But I have argued that a modest foundationalist position like Dogmatism amounts to a rejection of NM and with it, the idea that reception of new evidence is predictable to the agent. The Bayesian epistemologist cannot just *assume* NM in their objection of modest foundationalism. Such assumption amounts to a rejection of modest foundationalism from the get-go.

The above remarks are not intended as an objection to Bayesian epistemology. For one thing, a story has to be first told about the relationship between the numerical confidence levels of the Bayesian and the states of knowledge or justification. Instances of NM might very well be acceptable in certain applications of Bayesian epistemology, where one wants to e.g., offer a model of scientific confirmation. It does not follow that NM should be accepted in every context.[18]

White (2006) further discusses, following Vogel (2000), the *bootstrapping* objection to dogmatism. The bootstrapping problem is a serious issue that seems to generalize to any modest foundationalist theory of knowledge or justification (van Cleve 2003, Weisberg 2010). I believe that implicit diachronic introspection assumptions also play an essential role in the bootstrapping problem, and I defer this discussion to the next section of this dissertation.

3.5 Conclusions

I have argued that the No-Miracles principle plays an important role in the problem of perception. While skeptics assume that perceptual experiences need to be transparent to the subject in order to generate knowledge, prominent modest foundationalists reject this requirement. One important but relatively underdeveloped consequence of the good-case-versus-bad-case epistemological reasoning is that it makes perceptual experiences opaque to the epistemic subject. Some, like Williamson (2000), have argued that a response to the skeptic should involve the recognition that the subject's own mental states are not transparent to her. I have shown how this lack of transparency can be cashed in diachronic terms, as the inability to predict the *effect* of epistemic events.[19]

When this diachronic dimension is taken into account, modest foundationalism can be better understood. In particular, the apparent incompatibility with Bayesian epistemology can be explained

[17] For further discussion, see Weatherson (2007), Pryor (2013), Miller (2016).

[18] Of course, Bayesian epistemologists do offer arguments in favor of *the norm* of conditionalization (and so implicitly to NM), most prominently diachronic Dutch book arguments (Lewis 1999, van Fraassen 1984). The acceptance of NM is well motivated in certain Bayesian contexts.

[19] Unlike Williamson (2000: ch 7), the account I present here does not involve the failure of the KK principle. The dynamic (or diachronic) dimension of the epistemology of perception I focus on is logically independent of the static (or synchronic) dimension of one's higher-order mental states at a given time, which is the focus of the KK principle.

away by appealing to the underlying commitments of the theory.

It is no surprise that in both the formal Bayesian account of learning and standard accounts of epistemic logic, epistemic events are taken to be transparent to the idealized agent. Like other idealized assumptions, such as logical omniscience and full static introspection, this diachronic transparency assumption delivers simplicity and formal elegance.[20] It is also innocuous in many applications of these formal theories. It is nevertheless fruitful to understand under what conditions and theoretical commitments this diachronic transparency fails.[21]

[20] Logically, the connection between static and dynamic introspection runs deeper. In the possible world semantics of modal logic, assuming that the agent has *full static introspection* (i.e. both positive and negative introspection) amounts to the assumption that the set of worlds epistemically open to the agent is *constant* across the worlds the agent considers possible. Since that set is constant, the agent has no uncertainty about what they know. Similarly, assuming *full dynamic introspection* in the possible worlds semantics amounts to the assumption that the result of any given update is the same (i.e. constant) across the worlds in the model. Since the result of the update is constant across possible worlds, the agent has no uncertainty as to the effect of the update, and we can consider it transparent. For a logical framework for reasoning about failures of dynamic introspection, see Chapter 2 of dissertation, or Cohen (2020a). For more on the connection between the failure of static and dynamic introspection, see Chapter 5, or Cohen (2020b).

[21] I would like to thank Johan van Benthem, Ray Briggs, Krista Lawlor, the participants of the Stanford GSW, and two anonymous referees of the journal *Synthese* for many helpful comments, suggestions and corrections on earlier versions of this chapter.

Chapter 4

The bootstrapping problem and Perfect-Recall

Chapter abstract: Bootstrapping is a suspicious form of reasoning that seems to allow agents to gain knowledge about the reliability of their sources of information 'from thin air', without gathering independent evidence about those sources. The bootstrapping problem is the problem of explaining what is wrong with bootstrapping reasoning. I offer a solution to this problem by making explicit an implicit dynamic assumption which is made in bootstrapping reasoning, namely the Perfect-Recall principle. Despite its name, the Perfect-Recall principle is not necessarily about memory, but about the general ability to tell how we know what we know. In cases where we can't tell how we know what we know, the Perfect-Recall principle should be rejected, thus blocking the bootstrapping reasoning.

Suppose I look at a clock tower of a foreign village and see that it points to 12:00. I come to know that it is now noon. Does this give me reasons to believe that the clock is working properly? I might reason as follows: if by looking at the clock I get to know the time, then the clock must be working properly. If the clock is not working properly, then I would have not come to know the time. But I have looked at the clock and now I do know the time. Therefore, the clock must be working properly.

There is something intuitively wrong about the reasoning I just performed. It seems that in order to get information about the workings of the clock tower, I must gain further independent evidence about it. My belief that the clock is working properly cannot be solely *based* on the belief I formed by looking at the clock; that just does not seem to be enough.

The reasoning procedure described above is an instance of what is known as *bootstrapping reasoning.* I *bootstrap* my way into believing that the clock is working properly, since, it seems, I do not use independent epistemic resources to form this belief, even though these resources seem essential. The *bootstrapping problem*, as I understand it here, is the problem of explaining what is going wrong in bootstrapping reasoning. My aim in this chapter is to make explicit one implicit assumption that seemed to go unnoticed in many instances of bootstrapping reasoning, thus helping to address the bootstrapping problem.[1]

Here is a brief and incomplete sketch of the recent history of the bootstrapping problem, demonstrating its scope and centrality in contemporary epistemology. Fumerton (1995) and Vogel (2000, 2008) have argued that the bootstrapping problem is a serious problem for reliabilist theories of knowledge and justification. Cohen (2002) has argued that the problem extends to any theory of knowledge that allows for *basic knowledge*, i.e. the possibility of getting knowledge from a source of information without knowing that the source is reliable. This label applies to most externalist theories of knowledge, evidence and justification. Cohen uses the phrase *easy knowledge* to describe the "knowledge" obtained from bootstrapping. White (2006) has argued that bootstrapping is a problem for *Dogmatism* (Pryor 2000, 2004), the position that one can gain prima-facie justification from perceptual appearances without being justified that these appearances are veridical. Weisberg (2010) argues that the problem goes beyond dogmatist and basic knowledge theories. The problem is quite widespread.[2]

There are many versions of the bootstrapping problem, and it is not immediately obvious that all versions involve the same core phenomenon.[3] Vogel's original formulation of the bootstrapping problem involved some form of enumerative inductive reasoning to achieve the bootstrapping conclusion. Inductive knowledge has of course its own set of philosophical problems, and, following Titelbaum's (2010) insights, I believe that separating inductive matters from our discussion of bootstrapping can highlight core issues exclusive to the latter problem. I therefore focus on a version of the bootstrapping problem, adapted from Titelbaum (2010), that does not involve induction. I will exclusively address the issue of bootstrapping reasoning that purports to result in knowledge. Obtaining justified beliefs or increased levels of confidence by bootstrapping reasoning will not be discussed.

This chapter is structured as follows. In section 4.1, I present the bootstrapping example that will be at the focus of this chapter. In section 4.2, I present and criticize one common formalization of the bootstrapping reasoning. In section 4.2, I offer a different, dynamic formulation of the bootstrapping reasoning. I argue that this formulation reveals that bootstrapping reasoning implicitly assumes the dynamic *Perfect-Recall* principle (among other assumptions), and that there are good independent

[1] For a survey of proposed answers to the bootstrapping problem, see Weisberg (2012).

[2] See also van Cleve (2003), Goldman and Beddor (2016), Pryor (2013). Bootstrapping reasoning also plays a role in the epistemology of peer disagreement (Elga 2006, Kelly 2010).

[3] See e.g. the worries expressed in Titelbaum (2010) and the distinction made in Hasan and Fumerton (2018) between two types of bootstrapping.

reasons to reject this principle in the bootstrapping context, reasons that have nothing to do with memory or forgetfulness. I argue that assuming the *Perfect-Recall* principle amounts to assuming that the epistemic subject always knows *how* they come to know what they know. More precisely, according to *Perfect-Recall*, the agent is never uncertain as to the question *what exactly is the epistemic event that brought me to my current epistemic state?* Under my reconstruction, the *Perfect-Recall* principle is incompatible with such uncertainty. But since such uncertainty is plausible in bootstrapping scenarios, we have reasons to reject the latter principle, thus blocking the overall bootstrapping argument. Section 4.4 concludes.

4.1 The example

Here is the non-inductive bootstrapping example that I will analyze. This version of the problem is adapted from the non-inductive version of the bootstrapping problem from Titelbaum (2010). Noor sees that a clock tower is pointing to 12:00. Noor has the prior knowledge that the clock mechanism is always reliable or always anti-reliable. By anti-reliable clock mechanism we mean a clock mechanism that never tracks the right time. Since the clock mechanism is in fact reliable, and we assume the possibility of basic knowledge, we conclude that by looking at the clock Noor comes to know that the time is 12:00. Noor also knows that if it is the case that just by looking at a normal clock she is able to come to know the time, it must mean that its mechanism is working properly. After looking at the clock, she knows both that the clock is pointing to 12:00 and that it is in fact 12:00, so she concludes—thus coming to know—that the clock mechanism is always reliable, just by observing the clock once.

Theorists that reject the basic knowledge assumption can give a simple explanation as to what is wrong in the above example: since Noor does not antecedently know that the clock mechanism is working properly, i.e. she does not know that it is a reliable source of information, she cannot come to know the time just by looking at the clock. Since she does not have knowledge of the time, she cannot start applying the bootstrapping reasoning. This maneuver is not available to theorists that accept the possibility of basic knowledge. In this chapter I am going to ignore this maneuver and focus on the bootstrapping problem for basic knowledge theorists.

In inductive versions of the bootstrapping problem (like that of Vogel (2000)), the agent does not start with the assumption that the source of information (here, the clock mechanism) is always reliable or always anti-reliable. Instead, we consider multiple situations, in each of which the agent notices a correlation between the information given to them by the source and their current state of knowledge. By noticing this correlation n many times, the agent eventually comes to know, by inductive reasoning, that the source of information is in general reliable. As Titelbaum (2010) notes, by adding the assumption that the agent knows that the source is always reliable or always anti-reliable, we can set aside the inductive part of Vogel's bootstrapping argument.[4]

I am going to offer two formal explications to my bootstrapping example. The first follows the way Vogel reconstructs the problem. The second is my own. I will argue that the first formulation is lacking and that the second reveals an implicit assumption in the bootstrapping reasoning.

4.2 A standard formulation of bootstrapping

Like Vogel (2000), we will use the formal language of epistemic logic to analyze the bootstrapping reasoning. We use the formula $K\varphi$ to represent Noor's knowledge that φ. Denote with $C(12:00)$ the proposition that the clock is pointing at 12:00, and with $T(12:00)$ the proposition that the time is actually 12:00. Further denote with r the proposition that the clock mechanism is always reliable. We can explicate the assumptions of the bootstrapping reasoning with the following three premises:

(1) $K(C(12:00))$

(2) $K(T(12:00))$

(3) $K(C(12:00) \wedge T(12:00) \rightarrow r)$

Premise (1) states that Noor knows that the clock is pointing at 12:00. Premise (2) states that Noor knows that the time is 12:00. Premise (3) states that Noor knows that if the clock says it is 12:00 and the time is 12:00, then the clock mechanism is accurate in this occasion. (3) is supposedly supported by the fact that Noor knows that the clock mechanism is always reliable or anti-reliable (I will soon argue that this is inaccurate). Assuming the closure of knowledge, sentence (4) follows:

(4) $K(C(12:00) \wedge T(12:00))$

(4) says that Noor knows that the clock is pointing at 12:00 and the time is 12:00. It follows from (1) and (2) given closure. Given closure again, we get (5) from (3) and (4):

(5) $K(r)$

i.e. Noor knows that the clock mechanism is reliable.

Here are a few controversial assumptions that were made in the last paragraphs and which I will not focus on in my analysis of the bootstrapping problem. First, the above analysis assumes an unrestricted form of the closure of knowledge. The closure of knowledge is a controversial principle in general, and in particular in the context of bootstrapping (Cohen 2002, Luper 2018). Moreover, I will not discuss the assumption that warrant transmits across competent deductive inferences, which

[4] This non-inductive version is nevertheless different from some bootstrapping arguments in which the bootstrapping conclusion is what we may call, following Dretske (2005), *heavyweight* conclusions, like *there is an external world.* This type of problem is sometimes called the *easy knowledge* problem, following Cohen (2002). I will not discuss the connection between Vogel style bootstrapping cases and those cases that involve heavyweight propositions (for more on this distinction between the two types of bootstrapping, see Hasan and Fumerton (2016)).

has been linked to bootstrapping reasoning (see Moretti and Tommaso 2018 for an overview). I will not discuss those issues not because I think they are unimportant, rather because my aim is to focus on a different implicit assumption that occurs in bootstrapping reasoning. I am not claiming that there is a unique mistake made in bootstrapping reasoning; for all I know there are many.

Instead, I will focus on the following problem in the derivation of (5) from (1)-(3). The major issue is that the bootstrapping reasoning has certain dynamic (or diachronic) features that the formulas presented in (1) to (5) ignore. Noor's knowledge that the time is 12:00 is the *result of* (alternatively, *based on*) the epistemic event of looking at the clock tower and receiving the information that it points to 12:00. So the knowledge expressed in (2) represents the *posterior* knowledge Noor has *after* looking at the clock. On the other hand, the knowledge expressed in (3) does not seem to be a knowledge state that results from observing the clock. (3) is better understood as a kind of *prior, background* knowledge that Noor has. But (3) does not accurately represents Noor's prior knowledge of the situation, as I will now argue.

First of all, it is not clear that $C(12:00) \wedge T(12:00) \rightarrow r$ correctly describes the situation. To see why, consider again the version of our story where the clock mechanism is in fact anti-reliable, but the clock face is showing the right time, because of some deviant causal chain. Even if the clock mechanism is anti-reliable, it is still possible that the clock face is showing the right time. In such deviant scenario, at 12:00 o'clock, it is both true that $C(12:00) \wedge T(12:00)$, but if the clock mechanism is anti-reliable, then r is false. Therefore, the sentence $C(12:00) \wedge T(12:00) \rightarrow r$ does not seem to entirely capture the assumptions we make in the example, and its therefore not clear that it is known by Noor. Second, the sentence $K(C(12:00) \wedge T(12:00) \rightarrow r)$ does not specify that Noor knows the time as *a result* of looking at a normal clock tower. Complicating the deviant situation even further, it is possible that the clock mechanism is in fact anti-reliable ($\neg r$), Noor knows that the clock face is showing 12:00 by looking at the clock ($K(C(12:00))$), and she knows the time is 12:00 ($K(T(12:00))$) based on entirely different source of information. Maybe she learned the time by checking her phone, which she knows to be a reliable source of information. Obviously this scenario is not what we have in mind; we want our explication of the example to rule out such cases.

A better way to describe Noor's background knowledge is the following: Noor knows that if *as a result of* the event of looking at a normal clock, she comes to know the actual time, the clock mechanism must be reliable. This formulation of her background knowledge is not susceptible to counterexamples like that of the last paragraph. In the case where the broken clock points to 12:00 and Noor comes to know that the time is 12:00 by looking at her phone it is *not true* that Noor comes to know the time by (or as a result of) looking at the clock tower. The antecedent of her background knowledge is false, so the background knowledge is not threatened. Thus, the dynamic aspect of the *dependency* between the event of looking at the clock tower and the epistemic state of knowing the time is crucial for our understanding of the bootstrapping scenario. Since this dimension of the bootstrapping reasoning is completely absent from the formal representation of the problem given in (1)-(5), I suggest moving to a richer formalization that allows us to accommodate it. This is the aim of the next section.

4.3 Bootstrapping and dynamic introspection

In order to formalize the dynamic aspects of our bootstrapping reasoning I will extend the simple epistemic logic used in the last section with *event* operators, thus essentially working in (a simplified version of) *dynamic epistemic logic*.[5] The most basic axioms of dynamic epistemic logic are those that describe the interaction between epistemic events and epistemic states, knowledge in particular. I will call these interaction axioms *dynamic introspection principles*, because they can also be understood as encoding what the agent knows about the effects and history of the relevant epistemic events.

To reason about epistemic events like *looking at a normal clock tower that points to 12:00* we add a new type of propositional operator, E, beyond the epistemic K operator. The sentence $E\varphi$ reads 'as a result of event E, φ is the case'. To see how event operators play in our formalization of the bootstrapping reasoning, note that we can reformulate premise (2) from before, stating that as a result of looking at a clock that points to 12:00 Noor knows that the time is 12:00, as:

(6) $[C(12:00)]K(T(12:00))$

where the event E in this instance is $[C(12:00)]$, standing for the event of looking at a normal clock tower that points to 12:00. We read $[C(12:00)]K(T(12:00))$ as stating 'As a result of the event of looking at a normal clock that points to 12:00, Noor knows that the time is 12:00'. (6) explicitly describes the dependency between the event in question and the knowledge state of Noor.

With event operators we are also in a much better position to formally represent the background knowledge Noor has about accurate clocks. Recall that we informally expressed Noor's background knowledge as: Noor knows that if *as a result* of looking at the clock, she comes to know the actual time, the clock mechanism is reliable. This can be formally represented with the following premise:

(7) $K[C(12:00)](K(T(12:00)) \rightarrow r)$.

(7) says that Noor knows that as a result of looking at a normal clock tower that points to 12:00, if she comes to know that the time is 12:00, then the clock mechanism must be reliable in general. The background knowledge expressed in (7) is knowledge about the dependency between the event $[C(12:00)]$ and the conditional $K(T(12:00)) \rightarrow r$. There is also a 'syntactic' sense to call (7) background knowledge: the outer K is not scoped by any event operator, thus representing the

[5] See Baltag and Renne (2018) for a proper introduction to dynamic epistemic logic. See Chapter 2 for a version of dynamic epistemic logic that can violate the PR principle, as needed here.

initial, or prior, knowledge Noor has about the event $[C(12:00)]$. The truth of (7) depends of course on our assumption that Noor knows that the clock is either reliable or anti-reliable.

Note that (7) is *not* equivalent to the claim that Noor has the prior knowledge that the clock is reliable. (7) expresses the much weaker prior knowledge that *if* the event of looking at the clock results in knowledge of the time, *then* the clock mechanism is reliable. (7) is consistent with the assumption that Noor does not have the prior knowledge that the clock is reliable, and so with the assumption that Noor's knowledge of the time after looking at the clock counts as *basic knowledge.* We do not assume that the knowledge expressed in (7) is needed in order for Noor to come to know the time. In other words, we assume that (6) and (7) are independent: Noor would have come to know the time (assumption (6)) even if (7) were false. The truth of (7) does not change the status of Noor's knowledge of the time as basic.

Our dynamic reconstruction of the bootstrapping reasoning thus far includes two premises, namely:

(6) $[C(12:00)]K(T(12:00))$

(7) $K[C(12:00)](K(T(12:00)) \to r)$.

The question now is what else we need to assume in order to derive the bootstrapping conclusion $[C(12:00)]K(r)$, i.e. that as a result of looking at the clock, Noor is in a position to know that the clock mechanism is working properly. Informally, the argument goes as follows: (7) states that Noor knows that if as a result of looking at the clock, she comes to know the time, then the clock mechanism must be reliable (this is her background knowledge). (6) states that as a result of looking at the clock, Noor actually comes to know the time. Thus, after looking at the clock, Noor should be in a position to use her background knowledge (expressed in (7)) to conclude that the clock is reliable. This is the bootstrapping conclusion. The rest of this section is devoted to making explicit the implicit assumptions that are made in this informal argument.

The dynamic introspection principle that will be relevant for my analysis of the bootstrapping reasoning is:

Perfect-Recall (PR): $KE\varphi \to EK\varphi$

Perfect-Recall states that if the subject is *antecedently* in a position to know that event E results in φ being the case, then once event E happens, the subject knows φ.[6] We can think of the φ in the PR principle as the 'mark' of event E. The antecedent of PR states that the agent knows that φ is the mark of an E event. The consequent states that as a result of E, the agent knows the mark. Thus, according to PR, assuming that the agent knows the mark of an event, after the event the agent has no uncertainty as to which event just occurred. As the name suggests, it is

[6] See Chapter 2 for a logical overview of the principle. For an overview of PR and other dynamic introspection principles in their 'home ground' of game theory and epistemic logic, see van Benthem et al. (2009) van Benthem and Klein (2019), van Benthem (2012), Wang and Cao (2013).

natural to interpret this condition as characterizing the perfect memory of an agent: if, at first, the agent knows that event E results in φ, then the agent does not forget that later, once E actually happens.[7] As I will argue soon, however, failures of Perfect-Recall can occur even if no forgetfulness is assumed. An agent might have the *de-dicto* knowledge that event E is marked by a φ state, while having the *de-re* ignorance that *this* event is an E event. Such ignorance results in the failure of PR, as we shall soon see.

Perfect-Recall, I claim, is crucial for the bootstrapping reasoning. An instance of the Perfect-Recall principle, together with an instance of the KK principle and the closure of knowledge will allow us to get our bootstrapping conclusion from premises (6) and (7). Here is the derivation that shows how (followed by a detailed explanation):

(a) $K[C(12:00)](K(T(12:00)) \to r)$ premise (7)

(b) $K[C(12:00)](K(T(12:00)) \to r) \to [C(12:00)]K(K(T(12:00)) \to r)$ PR

(c) $[C(12:00)]K(K(T(12:00)) \to r)$ MP on (a) and (b)

(d) $[C(12:00)]K(T(12:00))$ premise (6)

(e) $[C(12:00)]KK(T(12:00))$ KK principle

(f) $[C(12:00)]K(r)$ closure of knowledge, from (c) and (e)

Derivation (a)-(f) is my proposed formulation of the bootstrapping reasoning. Let me explain it: (a) is just premise (7), the assumption that Noor has the background knowledge that if looking at the clock tower results in knowledge of the time, then the clock mechanism must be reliable. (b) is an instance of the Perfect-Recall principle $KE\varphi \to EK\varphi$, where $E = [C(12:00)]$ and $\varphi = K(T(12:00)) \to r$. (c) is obtained by (a) and (b) via modus ponens. (c) states that as a result of looking at the clock, Noor knows that if she knows the time, then the clock mechanism must be reliable. Supposedly, line (c), which is obtained by the Perfect-Recall assumption in line (b), states that Noor does not lose her background knowledge (line (a), premise (7)) after looking at the clock. This is not entirely accurate, as I will soon argue. Line (d) is premise (6), our *basic knowledge* assumption that Noor gains knowledge of the time by looking at the reliable clock. (e) is obtained by an instance of the KK principle, stating that as a result of looking at the clock Noor knows that she knows the time. Line (f) is our bootstrapping conclusion: Noor knows that the clock mechanism is reliable after looking at the clock. It is obtained by the closure of knowledge, given lines (c) and (e).

The derivation involves three substantial assumptions: the closure of knowledge, the KK principle and the Perfect-Recall principle. This is good, since, recall, our aim was to make explicit the implicit assumptions that underlie the informal bootstrapping reasoning. Much has been written

[7] See, e.g., Halpern (2004) for an analysis of the role of *Perfect-Recall* in the sleeping beauty debate, and, more generally, in Bayesian conditioning and van Fraassen's (1984) reflection principle.

about failures of closure and higher order knowledge, so my focus here will be on the Perfect-Recall assumption. There are good independent reasons to suspect that a principle like Perfect-Recall will fail in cases of bootstrapping that emerge from basic knowledge, and these reasons offer us an(other) answer to the question what goes wrong in bootstrapping reasoning.

Let us first clear out one bad reason for rejecting the Perfect-Recall assumption in bootstrapping reasoning, namely forgetfulness. Noor can know, before looking, that as a result of looking at the clock, if she knows the time, then the clock mechanism is reliable. She might forget that piece of information immediately after looking. If this is the case, then clearly line (b) of the derivation should be rejected. But this is a bad objection, since the bootstrapping reasoning I present here does not have anything to do with memory loss. For all we care, we can just assume that Noor has perfect memory when she looks at the clock and performs her bootstrapping reasoning.

Perfect-Recall can fail, however, even in cases where the subject has perfect memory. It can fail when there is a certain discrepancy between the agent's *de-dicto* knowledge about the effect of an epistemic event and their *de-re* knowledge of the epistemic event itself. This can happen because the agent's abilities to distinguish between distinct epistemic events might be imperfect. Recall that Perfect-Recall states that $KE\varphi \rightarrow EK\varphi$. Suppose that the antecedent holds, i.e. $KE\varphi$. This can be understood as a kind of *de-dicto* knowledge the agent has: the agent knows that *event E results in φ*. According to Perfect-Recall, it follows that $EK\varphi$, i.e. as a result of event E, the agent knows φ. On a natural interpretation, this step seems to assume that the agent has the *de-re* knowledge that the event that just happened is event E. If for all the agent knows, the event that just happened might be E' and not E, then even if the agent has the *de-dicto* knowledge that event E results in φ, the agent cannot conclude φ, since the agent does not know that E happened.[8]

When PR is stated explicitly, it might seem obvious to us that it cannot be unrestrictedly true. It is far from obvious, however, that we implicitly avoid every principle that we explicitly reject. Compared to epistemic-logic principles, like closure or KK, PR is relatively unknown, which makes it easier to miss when analyzing arguments like the bootstrapping problem. This is all the more important to remember if the dynamic dimension of the reasoning pattern is itself left implicit, as was done, for instance, in section 4.2.[9] This chapter is not primarily about rejecting PR; it is about highlighting the subtle role it plays in reasoning patterns like the bootstrapping reasoning.

Lois Lane's situation exemplifies a failure of Perfect-Recall that has nothing to do with memory. Let event E be the event of meeting Superman, and let φ be the proposition *Lois Lane is having a special day.* Then both $KE\varphi$ and $\neg EK\varphi$ are true, which is a counter example to the Perfect-Recall principle that $KE\varphi \rightarrow EK\varphi$. $KE\varphi$ is true because Lois Lane has the *de-dicto* knowledge that

[8] The *de-re de-dicto* distinction can be syntactically analyzed in terms of the scope order of a given quantifier and an intensional operator, like knowledge (see Nelson (2019)). Note that, syntactically speaking, *Perfect-Recall* allows one to switch the order of quantification between the event and knowledge operators. In this sense it is no surprise that the failure of *Perfect-Recall* could be cashed out in terms of the difference between *de-re* and *de-dicto* knowledge.

[9] Moreover, many formal systems that focus on dynamic aspects of reasoning, like standard dynamic epistemic logics and the Bayesian framework for update, tend to implicitly assume PR.

meeting Superman is a special event; $\neg EK\varphi$ is true because Lane is not able to identify Clark Kent as Superman, thus she does not know that a special event just occurred. No forgetfulness is involved on Lane's part, but Perfect-Recall fails.

A similar kind of ignorance can occur in Noor's clock example. I will first characterize it in schematic terms and then offer a more concrete (but somewhat outlandish) example. Suppose that there are two distinct events, E and E' s.t.: (i) both E and E' can result in knowledge that p according to a given basic knowledge theory, and (ii) the agent cannot distinguish between E and E'. E and E' can be two epistemic events that result from different reliable processes, both of which can produce knowledge that p, which the agent cannot distinguish. Now suppose that the agent has the *de-dicto* knowledge that event E results in ψ, i.e. $KE\psi$, and further suppose that event E' does not result in ψ. If event E actually occurs, then for all the agent knows, it was actually event E'. Thus, the agent cannot conclude that ψ is now the case, even though event E indeed happened. This is a failure of Perfect-Recall because both $KE\psi$ and $\neg EK\psi$ hold.

To show that such failure of Perfect-Recall is reasonable in Noor's clock example one has to consider two events E and E' that fit the schematic description given above. Let E be our event of looking at a normal clock tower, i.e. the event $[C(12:00)]$ we had from before. When conditions are right, the event $[C(12:00)]$ can generate knowledge of the time. Second, consider a somewhat stranger possible event. Consider the possibility that the village Noor is visiting has the tradition that for any anti-reliable clock mechanism in the village, there is a person inside that clock tower that makes sure that the hands of the clock show the right time. We can further suppose that the village takes this job very seriously, and so that the person inside the clock makes sure to reliably and accurately show the time when the clock mechanism is anti-reliable. Denote this person-inside-a-clock event $[C(12:00)]^*$. By assumption, the event $[C(12:00)]^*$ is different than the event $[C(12:00)]$, while both events can reliably produce knowledge of the time (albeit via very different processes). Moreover, Noor cannot, just by having a glimpse at the face of the clock, distinguish between event $[C(12:00)]$ and event $[C(12:00)]^*$.

Noor knows that if she is looking at a normal clock and comes to know the time, the clock mechanism must be reliable. Noor does not know that if she is looking at a strange village clock with a person inside it and comes to know the time, then the clock mechanism must be reliable. In other symbols, the formulas $K[C(12:00)](K(T(12:00)) \to r)$ (i.e. premise (7)) and $\neg K[C(12:00)]^*(K(T(12:00)) \to r)$ are true in our bootstrapping situation. Moreover, the bootstrapping situation is such that event $[C(12:00)]$ is the event that actually happens. But since for all Noor knows, it is event $[C(12:00)]^*$ that happened, she cannot conclude, after the event, that $K(T(12:00)) \to r$ is the case. For all she knows $K(T(12:00)) \land \neg r$ is the case, i.e., the clock mechanism is anti-reliable but she knows the time because there is a person inside the clock making sure to show the right time.

To see how this is related to our PR assumption (b) in the derivation, recall that (b) states that:

$K[C(12:00)](K(T(12:00)) \to r) \to [C(12:00)]K(K(T(12:00)) \to r)$.
If Noor cannot rule out that she came to know the time by the event of looking at a broken clock tower with a person inside it, the antecedent of (b) is going to be true while the consequent false. Noor can have the (*de-dicto*) knowledge that if she is looking at a normal clock, then if she comes to know the time, then the clock mechanism must be reliable. This is why the antecedent of (b) can be true. But since Noor cannot distinguish the event of looking at a normal clock and the event of looking at a strange village clock with a person inside it, after looking at the clock, she does not know which event has just happened. For all Noor knows, she gained knowledge of the time by a person inside a broken clock, not by a properly functioning clock. If this is the case, then for all she knows, she knows the time and the clock mechanism is anti-reliable. In other words, the consequent of (b) is false.

Generalizing from this example, in order to get counterexamples to the Perfect-Recall principle in bootstrapping cases one can take into consideration the ignorance the subject might have about *benevolent demons*. By *benevolent demons* I mean possible situations in which the subject gains knowledge as a result of an epistemic event which is different than the actual event that produced knowledge. I remain silent here as to whether these benevolent demons can actually produce knowledge.[10] We only need to assume that the bootstrapping subject cannot *rule out* the benevolent demon situation. This uncertainty is sufficient in order to block the application of the Perfect-Recall principle and with it the overall bootstrapping reasoning. Since Noor cannot rule out that she is in a benevolent demon situation (although she is actually not in one), i.e. she cannot rule out the possibility that she obtained knowledge of the time not by looking at a functioning clock, but by looking at a broken clock which is made sure to point to the right time by a person inside it, she cannot conclude anything about the functions of the clock mechanism.

Here is another way to put it. Noor might know the time, and she might even know that she knows the time. All of this does not imply that she knows *how* she knows the time. The question *whether* I know that p is different than the question *how* I know p. Consequently, the truth of the KK principle is a different matter from the truth of the Perfect-Recall principle.[11] Even if we assume KK, it does not mean that we assume that Noor knows how she came to know the time. If Noor cannot rule out the benevolent demon possibility, she does not know how she got to know the time. Given this uncertainty, Noor cannot bootstrap her way into the conclusion that the clock mechanism must be working properly. One implication of the Perfect-Recall principle is that it does not allow this kind of uncertainty. As far as uncertainty as to how I got my knowledge is reasonable in bootstrapping cases, it is reasonable to reject the Perfect-Recall principle.

To sum up: I have argued that the derivation (a)-(f) is better than derivation (1)-(5) in capturing the informal bootstrapping reasoning, because it is able to capture assumptions about how the

[10] See, e.g. Greco (2010) for a relevant discussion.

[11] The dynamic *Perfect-Recall* principle is logically independent from the static KK principle in epistemic logic, as Chapter 2 shows.

epistemic event in question *bases* (or *results in*) the corresponding epistemic states. The derivation (a)-(f) shows that the Perfect-Recall assumption is crucial in the bootstrapping reasoning. It might not be the only crucial assumption, and it might not be fallacious in every context. But there are good reasons to assume that the Perfect-Recall assumption should be rejected in cases of basic knowledge, according to basic knowledge theorists. Basic knowledge theories do not expect the agent to know that their sources of information are reliable. This implies that such theories allow the agent to be ignorant about the exact identity of their sources of information (as, clearly, the reliable source is distinct from the unreliable one, and both are epistemically open to the agent). It is thus natural for basic knowledge theories to allow for ignorance about the agent's source of information, which then forces the rejection of the Perfect-Recall principle.

With this in mind, the centrality of Perfect-Recall assumption in bootstrapping reasoning can be appreciated without the dynamic formalism. Take Vogel's (2000) original bootstrapping story, the case of Roxanne and the gas gauge. Roxanne does not initially know that her car's gas gauge is reliable. Each day, Roxanne sees that her gas gauge reads FULL, so (assuming basic knowledge) she comes to know that her gas tank is indeed full, as the gas gauge is in fact a reliable source of information. By noting the match between the status of the gas gauge and the status of the gas gauge n times, she bootstraps her way to the conclusion that the gas gauge is in general a reliable indicator of the gas tank.

This bootstrapping reasoning implicitly assumes Perfect-Recall: even if Roxanne knows that she knows that the tank is full, and even if she is able to get new knowledge by competent deductions from what she knows, in order to come to know that the gas gauge is working properly in any particular occasion, she has to know that her knowledge of the state of the gas tank *resulted from* the event of looking at a properly working gas gauge. If Roxanne cannot rule out a benevolent demon, then even if she knows that the gas tank is full, she might not know *how* she knows that. Maybe she suspects that her friend hacked her dashboard, making sure to show all the right indicators. With this uncertainty, she cannot learn anything *about* the gas gauge, even if she can learn *from* it. The failure of the Perfect-Recall principle captures this type of uncertainty. If in every particular occasion Roxanne cannot rule out a benevolent demon, she will not be able to conclude anything in general about the workings of the gas gauge by induction. My focus on a non-inductive version of the bootstrapping reasoning makes it easier to concentrate on the Perfect-Recall principle; but we can see that the latter principle also plays a crucial role in the inductive versions of the problem.

Like with evil demons, when it comes to benevolent demons, we can stretch our imagination as much as we want. According to one form of skepticism, we don't have knowledge of the external world because we don't know that our perceptual capacities are reliable. The basic knowledge theorist responds by claiming that we don't need to know that our perceptual capacities are reliable in order to get knowledge from them. On one extreme end of the basic knowledge position, we don't need to know anything about our sources of knowledge in order to gain knowledge from them. If this is the case, then for all we know it is by direct revelation that we have our knowledge of the external world, not by our perceptual capacities. Given this epistemological position, the skeptical worry about the reliability of our senses becomes irrelevant, as it *assumes* that our knowledge is the result of events of perceptual experiences, an assumption that the position can suspend judgment on.

Failures of Perfect-Recall do not have to be abnormal or outlandish. Given the right level of description, and assuming the possibility of basic knowledge, they can be generated quite easily. Here is a very mundane failure of Perfect-Recall: suppose Ann and Bob are the only two reporters in the local newspaper, and that they are both reliable sources of information. I read an article in the newspaper that does not mention the author. The article was in fact written by Ann, but for all I know it was written by Bob. Assuming basic knowledge, we can suppose that I come to know p as a result of reading the article (since my source of information was reliable). Let A be the event of *reading the article written by Ann*, and let B be the event of *reading the article written by Bob*. It is event A that actually happened, and so the event that resulted in me knowing p. But since I cannot distinguish between event A and event B, I do not know that it is event A that happened. Since I cannot rule out that I came to know p by an article written by Bob, I don't exactly know how I came to know what I know. As long as I do not *exactly* know how I know p, the epistemic possibility of *some* benevolent demon is open to me.

4.4 Conclusions

I have argued that the *Perfect-Recall* principle is implicitly assumed in bootstrapping reasoning. *Perfect-Recall* essentially amounts to the assumption that we know what epistemic event has brought us to our current epistemic state. But in cases where we can know *that* we know without knowing exactly *how* we know, then *Perfect-Recall* should be rejected.

My answer to the bootstrapping problem is not meant to override other existing solutions. My primary aim is not to show how to avoid a particular epistemic puzzle, but to show how to make *explicit* one of the *implicit* dynamic assumptions that we take for granted (without even noticing) when we reason about knowledge change. Examples that involve bootstrapping reasoning happen to be a perfect case study for such an endeavor. More generally however, it is worthwhile understanding how different epistemological positions interact with different dynamic-epistemic principles. At the very least, this will allow us to characterize certain epistemological disagreements more clearly.

Chapter 5

Inexact knowledge and dynamic introspection

Chapter abstract: Cases of inexact observations have been used extensively in the recent literature on higher-order evidence and higher-order knowledge. I argue that the received understanding of inexact observations is mistaken. Although it is convenient to assume that such cases can be modeled statically, they should be analyzed as dynamic cases that involve change of knowledge. Consequently, the underlying logic should be dynamic epistemic logic, not its static counterpart. When reasoning about inexact knowledge, it is easy to confuse the initial situation, the observation process, and the result of the observation; I analyze the three separately. This dynamic approach has far reaching implications: Williamson's influential argument against the KK principle loses its force, and new insights can be gained regarding synchronic and diachronic introspection principles.

5.1 Introduction

According to *externalist* theories of knowledge, the factors that make knowledge different from mere true belief might be external, and so inaccessible, to the epistemic subject. Externalist theories thus seem to be in tension with the introspective capacities of epistemic subjects. This tension plays a key role in Timothy Williamson's work on perceptual knowledge and the failure of the KK principle, according to which *S knows that p* entails *S knows that S knows that p* (Williamson 2000). The cases that motivate Williamson are cases of *inaccurate* or *inexact knowledge*, which emerge whenever we gain knowledge from our imperfect, often inaccurate, perceptual capacities.

Cases in which the KK principle fails, according to the Williamsonian picture that emerges, can give rise to more extreme cases in which our second-order epistemic attitudes radically differ from our first-order attitudes: it is possible to know p while be extremely confident, given one's evidence, that one does not know p (Williamson 2014). Given the right (better yet, wrong) evidential state, our first-order epistemic life might be completely foreign to us. This is Williamson's story.

This story has influenced the way epistemologists have recently approached the question of higher-order evidence: in what ways should our higher-order evidence relate to our lower-order evidence? (Christensen 2010). According to *modesty*, it is sometimes rational not to be fully confident with regards to the question "what should my level of confidence be?" Rejecting modesty does not seem like a privilege that fallible creatures like us enjoy. But finding the correct way to combine first- and second-order evidential attitudes has proven to be a non-trivial philosophical task. There seems to be, however, agreement in the literature that cases of inexact knowledge, and more generally inexact observations, are cases in which conflicts between first- and higher-order evidence emerge.[1]

The general Williamsonian story has also been used to argue against the possibility of common knowledge (Lederman 2017), Good's Theorem in Bayesian epistemology (Salow & Ahmed, 2017), the Stalnakerian picture of assertion at the foundation of formal semantics (Hawthorne & Magidor, 2009), and standard assumptions in the epistemology of indicative conditionals (Holguín 2019), to name just a few applications. The Williamsonian account of inexact knowledge has proven to be extremely influential. At the same time, the KK principle — the rejection of which lies at the heart of the Williamsonian story — keeps having its defenders.[2] The debate is far from dead.

Here I want to tell a different story about the tension between introspection and inexact observations. I argue that Williamson's understanding of scenarios of inexact perceptual knowledge is incomplete as it stands and, in particular, fails to show that the KK principle is false. On the contrary, inexact perceptual observations are compatible with the KK principle, once the logical mechanism with which agents *update* their knowledge after they make inexact observations is clarified. Situations of inexact learning are not static, they are situations in which change of knowledge occurs due to a perceptual event. Consequently, I argue, the underlying logic of inexact observations should be *dynamic epistemic logic*, not its static counterpart. Cases of inexact observations do *not* force a (synchronic) conflict between first- and higher-order knowledge. At the same time, I argue, when inexact observations occur, epistemic agents cannot know how *future* evidence will affect their knowledge state, even if they are fully (synchronically) introspective. Inexact observations show that epistemic agents do not always have dynamic (or diachronic) introspection.[3] This kind of diachronic uncertainty is worth further inquiry, especially in the context of externalist theories of knowledge and evidence. I argue that the real conflict is between inexact observations and dynamic introspection, not the KK principle.

[1] See Horowitz (2014), Elga (2013), Dorst (forthcoming), Lasonen-Aarnio (2014, 2015), Roush (2017).

[2] For recent defense of the KK thesis, see Das and Salow (2016), Dorst (forthcoming), Goodman and Salow (2018), Greco (2014a,2014b, 2015, 2017), Stalnaker (2015).

[3] This is true for both logical and probabilistic formulations of introspection. One can add *evidential probabilities* in the style of Williamson (2014) to the epistemic models I present here. It can then be shown that on such models the probabilistic *diachronic reflection principle* fails when inexact observations occur. This is analogous to Williamson's (2014) demonstration that the probabilistic (synchronic) *reflection principle* fails on Williamson's static models. See section 5.4 of this chapter.

To my knowledge, nearly all earlier critiques of Williamson's argument against the KK principle have ignored the dynamic nature of inexact observations.[4] It is thus valuable to examine that dynamic aspect thoroughly, which is my aim here. In the rest of this section, I present Williamson's own understanding of inexact knowledge. I explain Williamson's understanding and formulation of the *margin-for-error principle*, which is central to his way of thinking about inexact knowledge. In Section 5.2, I develop my dynamic account of inexact knowledge and present a natural way to syntactically enrich Williamson's original argument in a dynamic language. Given my alternative dynamic reconstruction, a tension between a *dynamic introspection principle* and my dynamic formulation of the margin-for-error principle arises. This offers an alternative explanation to the tension between inexact observations and introspection. Section 5.3 contains my novel semantics for inexact updates, which enrich the syntactic analysis of Section 5.2 and sketches a general account of the epistemology of inexact observations. In Section 5.4, I show how my semantics can be extended with evidential probabilities, which allow us to rigorously explore the connection between inexact observations and the synchronic and diachronic reflection principles. Section 5.5 reevaluates Williamson's margin-for-error principle. I argue that given my alternative dynamic explanation, Williamson's *static* formulation of the margin-for-error principle should be rejected. The margin-for-error principle is a principle about knowledge *obtained from an inexact observation*, not about knowledge *in general*; Williamson's account fails to make this distinction. My dynamic account is able to capture the nature of inexact observations, including the motivation behind the margin-for-error principle, while avoiding the problems that Williamson's original picture faces. Section 5.6 concludes.

5.1.1 Williamson's Argument

For the sake of familiarity, I use the unmarked clock example (Williamson 2014, Elga 2013) as my guiding example in this chapter. Since I will end up offering a general account of inexact observations, my analysis can be generalized to other similar examples. Here is the scenario: we have an analog clock that lacks marks for hours and minutes. The hands of the clock point to 12:17. Ann is looking at the clock from afar. Since Ann has normal human perceptual abilities, it is not the case that after looking at the clock Ann knows that the clock is pointing at 12:17. However, Ann does learn something from looking at the clock. Ann has a margin-for-error, s.t. if the minute hand is pointing to minute i, then for all Ann knows it points to $i \pm 1$. That is, Ann's margin-for-error is 1 minute. Thus, after looking at the clock, Ann knows the following disjunctive proposition: the clock is either pointing at 12:16, 12:17 or 12:18.

[4] For these non-dynamic critiques, see Mott (1998), Brueckner and Fiocco (2002), Neta and Rohrbaugh (2004), Conee (2005), Dutant (2007), Greco (2014a) Halpern (2008), Egre and Bonnay (2008, 2009, 2011), Sharon and Spectre (2008) and Stalnaker (2015). Egre and Bonnay use dynamic epistemic logic, but not in order to model the act of observation. The only exception is Baltag and van-Benthem (2018), who take a dynamic approach different than mine.

First we present Williamson's argument against the KK principle in this context of inexact knowledge (2000: Chapter 5). Epistemic modal logic will be used to analyze the argument. We read the modal sentence Kp_i as "Ann knows that the clock is pointing at $12:i$;" $\hat{K}\varphi$ is an abbreviation for $\neg K\neg\varphi$, and translates to "for all Ann knows, φ." The first premise in the argument against the KK principle is that after looking at the clock Ann knows that the time is not 12:00, i.e.
P1: $K\neg p_0$.
The second premise of Williamson's argument encodes Ann's knowledge of her own margin-for-error. By the description of the example, the following should be true: $p_{i+1} \to \hat{K}p_i$; if the clock is pointing at $12 : i + 1$, then for all Ann knows (or: Ann cannot rule out that) the clock is pointing at $12 : i$. This claim follows from Ann's imperfect eyesight, and Williamson assumes that Ann knows this. Hence the second premise is $K(p_{i+1} \to \hat{K}p_i)$, or equivalently
P2: $K(K\neg p_i \to \neg p_{i+1})$.[5]
The third premise of Williamson's argument is the KK principle, stated as
P3:$Kp \to KKp$,
which is assumed for reductio. Assuming multi-premise closure of knowledge,[6] the following derivation holds:

$K\neg p_0$	by **P1**
$KK\neg p_0$	by an instance of **P3**
$K(K\neg p_0 \to \neg p_1)$	by an instance of **P2**
$K\neg p_1$	by closure
...	
$K\neg p_i$	
$KK\neg p_i$	by an instance of **P3**
$K(K\neg p_i \to \neg p_{i+1})$	by an instance of **P2**
$K\neg p_{i+1}$	by closure
...	
$K\neg p_{17}$	

The last line is a contradiction, given the assumption that p_{17} is true and the factivity of knowledge. Note how the same reasoning pattern is *iterated* multiple times in the derivation. Williamson's conclusion is that **P3** is false, i.e. the KK principle fails to hold in general.

Williamson's commitment to the second premise of the argument is partly based on his externalist epistemology, which is exemplified with a commitment to the safety condition of knowledge. The latter condition requires that if you know φ, then you could not have been wrong in very similar cases. Knowledge entails an error free buffer zone. In the unmarked clock example, we take the i case and the $i + 1$ case as very similar. Thus, Ann's imperfect eyesight together with the safety

[5] We assume that the implication in **P2** is material.

[6] The closure of knowledge will not be the focus of this chapter, even if its failure can break Williamson's derivation. See (Williamson 2000: p. 117) for a defense of closure in this context.

condition of knowledge implies that $p_{i+1} \wedge K\neg p_i$ is impossible. Supposing $K\neg p_i$ and that Ann cannot visually discriminate between i and $i+1$, it follows that Ann would have wrongly believed $\neg p_i$ in the very similar case in which p_i was true — contrary to the safety condition. Since $p_{i+1} \wedge K\neg p_i$ is impossible, $p_{i+1} \rightarrow \hat{K}p_i$ follows, and since all of this can be concluded by Ann with some reflection, we may assume she knows it, hence **P2**.[7] I will later argue for the rejection of **P2** (in Section 5.5), but my criticism is about Williamson's implementation of the buffer zone intuition (in the form of **P2**), not about the philosophical idea itself.

The concept of safety — as well as the standard semantics of epistemic logic — is modal. It is therefore natural to model the above syntactic argument using Kripke semantics. Recall that $K\varphi$ is true in a world w iff all the worlds related to w by the epistemic indistinguishably relation (denoted with R) are worlds in which φ is true. Moreover, since knowledge is factive, i.e. the K modality validates the $K\varphi \rightarrow \varphi$ axiom, the R relation must be reflexive. Consider the model in Figure 5.1, where the R relation is represented graphically by the solid arrows:

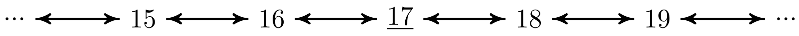

Figure 5.1: Williamson's intended model (reflexive arrows omitted, actual world underlined). The model satisfies **P1** and **P2** but not **P3**.

Note that at w_{17}, the world in which p_{17} is true (and the clock is pointing at 12:17), Ann knows $\neg p_0$, i.e. $w_{17} \models K\neg p_0$, as all worlds accessible from w_{17} with the R relation (i.e. w_{16}, w_{17}, w_{18}) are $\neg p_0$ worlds. Moreover, Ann knows the disjunction $p_{16} \vee p_{17} \vee p_{18}$ for the same reason. One can also check that $K\neg p_i \rightarrow \neg p_{i+1}$ is true at any world in the model: if Ann knows that p_i is not the case, then she is not located at a (very similar) $i+1$ world. Since $K\neg p_i \rightarrow \neg p_{i+1}$ is true everywhere, it is known by Ann, validating **P2**. **P3**, the KK principle, is false in the model. Example: in the actual world w_{17}, $K\neg p_{15}$ and $\neg KK\neg p_{15}$ are the case. For world w_{17} is accessible to w_{16} in which $\neg K\neg p_{15}$. Figure 5.1 models Williamson's conclusion of the unmarked clock argument: inexact knowledge without KK.[8]

KK fails, but Williamson's story moves on with a more ambitious twist. How confident should Ann be in the fact that she knows $p_{16} \vee p_{17} \vee p_{18}$? To answer that, we add *evidential probabilities* (Williamson 2000) to the model in Figure 5.1. Consider a uniform probability function P that assigns each world in the model a prior probability. So, in Figure 5.1, P assigns each world a prior weight of $\frac{1}{60}$, i.e. $P(w_i) = \frac{1}{60}$.

[7] If **P2** is based on the safety condition of knowledge, then one way of blocking Williamson's argument is by rejecting safety. See Neta and Rohrbaugh (2004) for this direction. But even if the safety condition is false in general, it seems like a very reasonable constraint in cases of inexact observations (Williamson 2008).

[8] I am not claiming that the model in Figure 5.1 is *the* intended model for the unmarked clock example, rather that it is a good enough simplified model. The model introduces many assumptions that go beyond **P1, P2** and **P3**. Further complications and considerations might be added, (see Williamson 2014) – but this is a good picture to start with.

To compute the evidential probability of the agent in a given world, one conditionalizes on the strongest proposition that one knows in that world. At w_i, the strongest proposition is identical with the set of all worlds accessible from w_i, denoted as $R(w_i)$. The evidential probability at world w_i, P_{w_i}, of some proposition X is $P_{w_i}(X) = P(X|R(w_i)) = \frac{P(X \cap R(w_i))}{P(R(w_i))}$. So for instance in Figure 5.1, the evidential probability of $p_{16} \vee p_{17} \vee p_{18}$ at world w_{17} is 1, as $P(p_{16} \vee p_{17} \vee p_{18}|R(w_{17})) = 1$. On the other hand, note that since the sentence $K(p_{16} \vee p_{17} \vee p_{18})$ is only true in world w_{17}, its evidential probability there is the significantly lower $\frac{1}{3}$, as $P(K(p_{16} \vee p_{17} \vee p_{18})|R(w_{17})) = \frac{1/60}{3/60} = \frac{1}{3}$. Ann knows $p_{16} \vee p_{17} \vee p_{18}$, but her confidence level in this fact are only $\frac{1}{3}$, so she is rather confident that she does not know $p_{16} \vee p_{17} \vee p_{18}$. Williamson (2014) shows how one can tweak this model to make Ann's confidence level in her own knowledge arbitrary low. Putting the knowledge operator aside, note the conflict that arises between Ann's first and second order evidential probabilities.[9] While she assigns probability 1 for the sentence $p_{16} \vee p_{17} \vee p_{18}$, she also assigns only $\frac{1}{3}$ probability to the sentence *my credence in* $p_{16} \vee p_{17} \vee p_{18}$ *is 1.0* (as it is only true in w_{17}). Doesn't this second order evidential state give her a good reason to lower her degree of confidence in the sentence $p_{16} \vee p_{17} \vee p_{18}$?

Williamson's and others seem to think that this *split* between our first and second order evidential states is a plausible state of affairs that is a part of our epistemic lives.[10] Others have argued that this cannot be the full picture and that there must be some principle of rationality that guides us in bridging our different pieces of evidence in such cases.[11] I want to argue that the unmarked clock case, and other cases based on inexact observations, are wrongly understood. The question is not about modeling inexact knowledge (or inexact evidence), it is about modeling the inexact *observation* that gets us from one knowledge state, before looking at the clock, to the second knowledge state, after the clock has been observed.

5.2 Introspection and Inexact observations

How should we model inexact observations? As we work with finite epistemic models and a coarse-grained conception of propositions, each observed proposition must have a clear boundary. But we want to model observations that lack this phenomenology of exact boundary. How should we do that? This *difficult* modeling question is not answered in Williamson's argument, as the act of observation itself is not part of the model. In Figure 5.1 we see the epistemic model after observation. But how did we get there?

Similar issues are reflected, to some degree, in Williamson's syntactic representation of his argument against KK. Williamson's use of the K operator is ambiguous between the knowledge state

[9] One might argue that Williamson's evidential probabilities assume the identity of knowledge and evidence (E=K), which many do not accept. Williamson addresses this issue in (2014, forthcoming). Models with evidential probabilities do not force E=K (although they assume the factivity of evidence), they are just taken as simple tool for investigating higher-order evidential states.

[10] See Williamson (2000, 2014) and Lasonen-Aarnio (2014, 2015).

[11] See e.g. Elga (2013) and Horowitz (2014).

before looking at the clock and the knowledge state *afterwards*. **P1** says that $K\neg p_0$, i.e. that Ann knows, *after looking*, that the minute hand is not pointing at 00. So we conclude that the K operator refers to Ann's knowledge after the observation. **P2** is not as clear. The inner K operator in $K(K\neg p_i \to \neg p_{i+1})$ seems to have the same meaning as the K operator in **P1**: knowledge after the observation. But the outer K operator can be read, prima facie, as representing Ann's knowledge *before* the observation. After all, according to Williamson, knowledge of margin of error is obtained by reflection, and this reflection can be done before looking at the clock.

It is crucial for Williamson's argument that at least some of the K operators are understood as knowledge after *the observation.* If a truth-telling Oracle tells Ann that the time is *not* between 12:00 and 12:16, then although **P1** remains true, **P2** becomes false. For then $K\neg p_{16} \wedge p_{17}$ is true, contrary to the margin-for-error principle. This is because in this context the K operator refers to *knowledge after accurate testimony*, and such knowledge state is not governed by a perceptual margin-for-error.[12] **P2** is not set in stone — it is highly dependent on the context of inexact observations. But none of these complications are reflected in the simple language of epistemic logic.

The above remarks, by themselves, do not count as objections to Williamson's argument. Given the kind of intended models Williamson is offering, it is rather clear that according to Williamson, all K operators should be understood as knowledge after the observation. However, a proponent of the argument should allow for this distinction to be made and incorporated into the argument. If the argument is robust, it should easily survive such modification.

So let us make this modification. Start by adding an *update* operator, a familiar addition from *dynamic epistemic logic* or DEL (Baltag & Renne 2016), to the formal language, representing the act of inexact observation.[13] Given any proposition e that can be observed, we add to the language the modal operator $[e]$, representing the observation that e. The modal operator $[e]$ is used to describe the result of the experience that e on Ann's epistemic state. As a rough approximation, we can read the formula $[e]\varphi$ as stating "as a result of Ann's veridical experience that e, φ is the case," or as "if Ann has the veridical experience that e, then φ", where the conditional is not understood as a materiel conditional.[14] By *veridical* we just mean that if we are in a not-e world, then the formula $[e]\varphi$ is vacuously true — our focus is on updating with veridical evidence that possibly generates knowledge.[15] $[e]$, as a 'box' type modal operator, has a 'diamond' dual $\langle e\rangle$, which is equivalent to

[12] A similar point is made by Sharon and Spectre (2008).

[13] Bonnay and Egre (2011) were first to consider the tools of DEL in order to analyze inexact knowledge. Their approach of using a non-standard semantics for the static base epistemic logic is very different from the one I develop here, as they do not use updates to model inexact observations. The very recent work of Baltag and van-Benthem (2018) uses what I call exact updates to analyze Williamson's inexact knowledge. The latter approach is quite different than mine as it does not try to offer an alternative or an explanation to the margin of error principle.

[14] Although it is possible, in this paper we will not develop the analysis of the DEL update operator as expressing a non-material conditional. See Icard and Holliday (2017) for an analysis of the relationship between the update operator of DEL and indicative conditionals. Even though the main function of the $[e]$ operator is to describe the effect of observing e on the agent's epistemic state, the formal syntax of DEL allows for expressions like $[e]p$ (where p is an atomic, non-epistemic formula). In those cases, there is clearly no dependency between observing e and p, and the conditional reading of the operator is more appropriate.

[15] The debate whether all evidence is veridical is not at issue here. Even if there is false evidence, such evidence does

$\neg[e]\neg$. The only difference between $\langle e\rangle\varphi$ and $[e]\varphi$ is in the treatment of non-veridical e: if e is false in the world of evaluation, then $\langle e\rangle\varphi$ is false, while $[e]\varphi$ is vacuously true. Put syntactically, the following is going to be a valid principle: $\langle\psi\rangle\varphi \leftrightarrow [\psi]\varphi \wedge \psi$. For ease of presentation, I will mainly use the box version of the update operator.

Conceptually, a key feature of the $[e]$ update operator is that it is an epistemic, but not necessarily a successful update. It is epistemic because it is used to model change of knowledge given true information (the agent's change of belief is not modeled in this analysis). At the same time, we are not assuming that it is epistemically successful: it is possible that after an update with true e, the agent does not come to know e. This feature will be important for modeling certain externalist intuitions later. Even in the case where e is true and the agent has the experience that e, we don't wish to assume that the agent automatically comes to know e, for the external environment can prohibit the agent from coming to know e (say because the agent is in an epistemically unsafe situation, or the source of information is unreliable in a given situation). Since $[e]$ is not assumed to be successful, it is hard to phrase in ordinary English; we cannot phrase $[e]\varphi$ with an expression like "as a result of learning e, φ is the case," as the latter assumes that the agent successfully comes to know e.

All discussed cases of inexact knowledge in the literature are cases in which the observation happens just *once* — there is no sequence of observations. Syntactically then, we will not nest the update operators inside other update operators. This simplicity leads to a clear way of distinguishing between knowledge *before the observation* and knowledge *after the observation*: every instance of the K operator inside the scope of the update operator represents the latter; all K operates outside the scope of an update operator represent the former. For instance, the first premise of Williamson's argument will be modified to $[e]K\neg p_0$: *after* Ann is making observation e, she knows that $\neg p_0$.[16]

Now, given our richer dynamic epistemic language (as opposed to its static fragment), one can locate principles that should be rejected on weak externalist considerations. Consider the following principle:

Dynamic Introspection (DI): $[p]Kp \rightarrow K[p]Kp$.[17]

not generate knowledge (at most it can generate false beliefs). Our focus is on knowledge update, so we can safely restrict our attention to veridical evidence and veridical update operators. We care to model knowledge update and not belief revision given any kind of (possibly false) piece of information. That being said, it is technically possible to extend the formal apparatus to accommodate such cases.

[16] Alternatively, one could use the framework of *epistemic temporal logic* (instead of dynamic epistemic logic) to represent the knowledge stages before and after the update with two distinct knowledge operators: K_0 and K_1. See van-Benthem et al. (2009) for details about the relationship between the two frameworks.

[17] In this chapter, I use the term dynamic introspection to describe the above dynamic principle so as to easily contrast it with the KK principle (static introspection). In the larger context of this dissertation, I note that the above principle is a consequence of any normal dynamic epistemic logic that accept the perfect-recall principle $K[\varphi]\psi \rightarrow [\varphi]K\psi$ (see Chapter 2). The consequent of DI is easily derivable as a theorem:

$\vdash p \rightarrow p$	propositional tautology
$\vdash [p]p$	atomic reduction axiom (see Chapter 2)
$\vdash K[p]p$	K-necessitation.
$\vdash [p]Kp$	perfect recall
$\vdash K[p]Kp$	K-necessitation.

(DI) says that if after the agent has a veridical experience of p, the agent knows p, then the agent knows (prior to the experience) that after a veridical experience that p, the agent will know p. Put differently: suppose that p is true, and that once the agent has the experience that p, she comes to know p. Then the agent knows that a veridical p experience generates knowledge. It seems that very weak forms of externalism generate counterexamples to (DI). After all, even if a piece of evidence p is true, it does not follow that it is also received safely (or reliably, or by the right causal connections). Consider the following example, formulated with a generic reliabilist language:

The Tree: There is a tree next to Bob. At t_1, Bob is having a veridical experience of a tree, and as a matter of fact, Bob's vision is reliable. Therefore, Bob knows that there is a tree in front of him at t_1. However, Bob does not know that his vision is reliable, and at t_0 he cannot conclude that the experience of a tree will result in him knowing that there is a tree in front of him. For all Bob knows, the experience of a tree might not be generated in a reliable fashion.

Note then that according to the example, we have (1) $[tree]K(tree)$ — after the veridical experience of $tree$, Bob knows $tree$, and (2) $\neg K[tree]K(tree)$ — Bob does not know, before the experience, that having the experience of $tree$ will result in knowledge of $tree$, since for all Bob knows, his perceptual faculties are unreliable. The conjunction of (1) and (2) provides a counterexample to (DI). I therefore take it as uncontroversial that generic forms of externalism are committed to the failure of (DI).

It is also worth remarking that (DI), like the KK principle, is a kind of introspection principle, where the consequent iterates a K operator on a condition in the antecedent. But (DI) and KK have different logical origins, as the former is dynamic while the latter is static. As a matter of fact, (DI) is a validity in standard forms of DEL, even if the underlining static epistemic logic is the modal logic T, i.e. a logic that invalidates the KK principle. In the logical picture that I am going to present, (DI) is false, while the KK principle remains true. Thus, the two principles are logically independent.[18]

All of this is related to Williamson's argument. I will now show that (DI) is incompatible with a plausible *dynamic* presentation of the unmarked clock example. Thus, Williamson's inexact knowledge argument could be interpreted as a reductio argument against (DI), once Williamson's premises are understood dynamically.

The problematic step in the Williamsonian derivation from the introductory section is the move from Ann's knowledge that $\neg p_{15}$ to her 'knowledge' that $\neg p_{16}$. The latter step is defective as we assumed that Ann knows $p_{16} \vee p_{17} \vee p_{18}$, but not more. In the derivation, we make this step by using $K(K\neg p_{15} \to \neg p_{16})$, an instance of **P2**. With the update operators, I will attempt to explain where the argument goes wrong. What follows is my reconstruction of the problematic part of the unmarked clock argument.

[18] It is thus also an independent question whether the Tree example is compatible with the KK principle. I leave this question aside. My point is that the example is clearly not compatible with Dynamic Introspection.

Fix e to be the proposition $p_{16} \vee p_{17} \vee p_{18}$. My **main assumption** is $[e]Ke$: after observing e, Ann knows e. It is an analytic truth that $e \rightarrow \neg p_{15}$, so we can assume that it is known by Ann before and after the observation, and known to be so. We thus have $K[e]K(e \rightarrow \neg p_{15})$; call this my auxiliary assumption. Given the closure of knowledge, the closure (or distribution) of the update operator, and our main and auxiliary assumptions we can deduce

$\mathbf{P1'}$: $[e]K\neg p_{15}$

(see the appendix for the full deduction). This is the first premise in *our* reconstruction of Williamson's argument. Next, we wish to translate Williamson's knowledge of margin-for-error assumption (his premise **P2**) into the dynamic language. I propose the following:

$\mathbf{P2'}$: $K[e](K\neg p_{15} \rightarrow \neg p_{16})$.

This is the dynamic version of Williamson's **P2**: Ann has the prior knowledge that after observing e, it is going to be the case that: if she knows $\neg p_{15}$, then p_{16} cannot be true, given her margin-for-error. For if the evidence e is strong enough to generate knowledge that excludes the p_{15} possibility, then, given Ann's observational margin-for-error, it could not have been generated in world w_{16}. Note that the addition of the dynamic operator allows us to distinguish knowledge before and after the observation. The inner K is scoped by $[e]$, thus representing knowledge after the observation that e; the outer K is not scoped by the operator, representing knowledge attained before the observation. $\mathbf{P2'}$ gives us a more fine-grained description of the situation at hand.

Recall that we assumed that the strongest proposition Ann learned from the observation is e. In Williamson's original derivation, we can derive $K\neg p_{16}$, which implies that Ann knows more than e, a contradiction to what we assume. Similarly, with our assumptions we can derive that Ann knows that after observing e, $\neg p_{16}$ is the case. This contradicts what we assumed, as we assumed that Ann does not know that $\neg p_{16}$ after observing e (we started with the assumption that the strongest proposition known to Ann after the observation is $p_{16} \vee p_{17} \vee p_{18}$). Thus, the **main assumption**, premises $\mathbf{P1'}$, $\mathbf{P2'}$ and (DI) lead to an absurdity, and I conclude that (DI) should be rejected.

I reserve the full formal derivation to the Appendix, and sketch it here: the **Main assumption** states that $[e]Ke$; together with (DI) and the auxiliary assumption, it follows that $K[e]K\neg p_{15}$. $\mathbf{P2'}$ states that $K[e](K\neg p_{15} \rightarrow \neg p_{16})$; under the assumption that the update operator distributes over implication, it follows that $K([e]K\neg p_{15} \rightarrow [e]\neg p_{16})$. By the closure condition of knowledge, it follows that $K[e]\neg p_{16}$ from the above two conclusions. In the form of a derivation:

(1) $[e]Ke$ **Main assumption**

(2) $K[e]Ke$ by (DI) on (1)

(3) $K[e]K(e \rightarrow \neg p_{15})$ auxiliary assumption

(4) $K[e]K\neg p_{15}$ closure (see Appendix)

(5) $K[e](K\neg p_{15} \rightarrow \neg p_{16})$. $\mathbf{P2'}$

(6) $K[e]\neg p_{16}$ from (2) and (3), by closure (see Appendix).

Line (6) contradicts the assumption that the strongest proposition that Ann learns from observing e

is e, as it states that Ann knows that observing e implies something stronger, namely $\neg p_{16}$. Crucially, note that we have derived a conclusion contrary to our initial assumption *without* the use of the KK principle.

The most controversial principle in the above paragraph is (DI). Much less controversial is the assumption that simple analytic truths like $e \rightarrow \neg p_{15}$ are known, and known to remain known after any veridical observation (i.e. $K[e]K(e \rightarrow \neg p_{15})$, the auxiliary assumption). We also assume that the update operator $[\cdot]$ distributes over implication: if after the φ update, $\alpha \rightarrow \beta$ is true, and after the φ update, α is true, then after the φ update, β is true. Assuming weak externalist tendencies, which are in tension with (DI) anyway, rejecting (DI) seems like the most plausible response.

Upshot: Williamson's formal argument does not distinguish between knowledge before and after the inexact observation. Once this distinction is made, we can enrich Williamson's argument with dynamic operators. When we do so, we see that the dynamic premises conflict with the dynamic principle I called (DI). This conflict offers an explanation to the tension between inexact observations and introspection. This dynamic explanation is different from Williamson's static explanation, as it involves different types of introspection principles. Note, however, that the alternative dynamic explanation is by itself not in conflict with Williamson's static argument. After all, the dynamic language is an *extension* of the static language; the two are perfectly compatible. It is hence fair to ask whether Williamson's static premises are true after the act of observation, and so whether inexact observations are in conflict with *both* static and dynamic introspection. In section 5.5 I will return to this question. Before I do that, I offer an account of inexact observations as a special kind of update, an *inexact update*.

5.3 A Semantic Perspective

While the last section has focused on the syntactic argument, here I present my novel semantics for inexact observation. The goal is to show that the KK principle and formulas like the **main assumption**, $\mathbf{P1'}$ and $\mathbf{P2'}$ are mutually compatible with the failure of (DI). A natural way to argue for compatibility of a set of assumptions is by invoking a model, which is what I do here. The model will also explain my novel approach to inexact observations as *inexact updates*.

The main challenge is to model the epistemic effect of inexact observations. To do so, we work with two models: the *initial model* (the situation before the observation has taken place), and the *updated model* (the situation after the observation). The semantic clause of the update operator will tell us how to compute the updated model from the initial model.[19]

The semantics of the K operator remains the same: $K\varphi$ is true iff φ is true at all epistemically

[19] In this chapter, I use a simple semantic system to model inexact, or opaque, updates. The more general semantic system detailed in Chapter 2 can be used to model the updates used here. In particular, the observation can e can be represented with the program $(?p_{17}; !(p_{16} \vee p_{17} \vee p_{18})) \cup (?p_{16}?; !(p_{15} \vee p_{16} \vee p_{17} \vee p_{18})) \cup (?p_{18}?; !(p_{16} \vee p_{17} \vee p_{18} \vee p_{19}))$. This is program has the same effect as the updates appearing in Figures 5.3-5.5

accessible worlds. The basic idea behind the semantics of the update operators in DEL is the familiar Stalnakerian notion of update: updating with proposition P has the effect of eliminating all the not-P worlds from the initial model. This is a good enough semantics for modeling exact updates, like learning from reliable testimony, but we will need to tweak it to accommodate inexactness.

I assume that in the initial model (before looking), Ann does not know anything about the position of the minute hand of the clock: $\neg Kp_i$, for any p_i. The principle that $Kp_i \to KKp_i$ thus follows vacuously. Moreover, we can assume that Ann knows that she does not know the position of the minute hand. Ann has no reason to think that she knows the time, and no deception is assumed in the example: hence $K\neg Kp_i$ holds for any i, and so $\neg Kp_i \to K\neg Kp_i$ follows as well. In order to capture these assumptions, we can let the initial model be an **S5** model in which all worlds are epistemically connected: Rw_iw_j for any i and j.[20]

Now, we add to the initial model another relation, the *perceptual inexactness* relation P, that intuitively specifies which worlds will be perceptually indistinguishable *during* an observation.[21] Unlike Ann's initial epistemic state, perceptual indistinguishably depends on which state is actual. If the actual state is i then, on observation, Ann cannot perceptually distinguish it from $i + 1$ and $i - 1$, according to the informal story about her margin-for-error. In other words, we have that for all w_i, Pw_iw_{i+1} and Pw_iw_{i-1}.[22] The *initial* epistemic model can be thus depicted as in Figure 5.2.

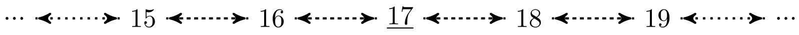

Figure 5.2: The initial model M according to my story (before Ann looks at the clock). The epistemic indistinguishably relation is the universal S5 relation, which is not depicted. The dashed arrows represent the relation P, of perceptual inexactness.

The model in Figure 5.2, unlike that of Figure 5.1, contains two relations: the universal epistemic R relation, connecting all worlds together (not depicted in the figure), and the perceptual inexactness relation P, depicted with the dashed lines.

What happens to the initial model when Ann has the e ($= p_{16} \vee p_{17} \vee p_{18}$) experience when looking at the clock at the actual world w_{17}? The idea is that we eliminate all the not-e worlds from the model, unless these worlds are perceptually indistinguishable from w_{17} (according to P). In general, the updated model, resulting from observing φ at w, has the following key property: all worlds in it are either (1) worlds in which φ is true, or (2) worlds that are perceptually indistinguishable from w. Put formally, the semantic satisfaction clause for the update operator is:

[20] Since the initial the model is an S5 model, positive and negative introspection hold for *any* φ, not just for the p_i's. Since what we care about is Ann's knowledge of the clock (i.e. about p_i), this idealization seems harmless.

[21] See Halpern (2008) for different approach that uses two relations in order to analyze related cases of perpetual vagueness. Dutant (2007) builds on Halpern's approach and provides an infallibilist critique of Williamson's margin-for-error argument.

[22] We could complicate the structure of the P relation to allow for varying margin-for-error the same way Williamson is varying his R relation in (Williamson 2014).

- $M, w \models [\varphi]\psi \quad \Leftrightarrow \quad \text{if } M, w \models \varphi \quad \text{then} \quad M_{\varphi,w}, w \models \psi$.

The antecedent on the right hand side of the equivalence guarantees that when φ is false, the expression $[\varphi]\psi$ is vacuously true. The consequent checks that ψ is true in the updated model $M_{\varphi,w}$, which is formally defined as $M_{\varphi,w} = (W', R', P', V')$:

$W' = \{v \in W \mid Pwv \;\; OR \;\; M, v \models \varphi\} = \{v \in W \mid Pwv\} \cup \{v \in W \mid M, v \models \varphi\}$

$R' = R \cap W'^2$

$P' = P \cap W'^2$

$V' = V$.

W', the set of worlds obtained by observing φ at w, is just the *union* of the φ worlds with the perceptually indistinguishable worlds. Note that if we wish to model the special case in which updates are exact, we just need to set the relation P to be empty. On such frames, the update operator behaves the same as in standard DEL.

The novelty of the above semantics lies in the fact that updating with the same φ at different worlds results in different updated models.[23] The leading motivation behind the semantics is based on a familiar intuition in epistemology: in the *good case*, veridical evidence generates more knowledge than in the *bad case*. Read, for instance, the *good case* as the cases where our perceptual capacities are reliable, and the *bad case* as the case where our capacities are not as reliable. Further assume that φ is a true piece of information (φ is true at the world of evaluation). In the good case, observing φ leads to knowing φ. In the bad case, observing φ leads to knowing a weaker proposition, say ψ. In the worst case imaginable, i.e. the extreme skepticism scenario, the veridical φ experience leads to no new knowledge at all (i.e. ψ is $\top$).

Let's consider the concrete example of the unmarked clock: updating with e at w_{17}. The result is in Figure 5.3:

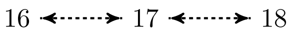

Figure 5.3: $M_{e,w_{17}}$ The result of updating with e $(= p_{16} \vee p_{17} \vee p_{18})$ at world 17.

Compare that to updating with the same e at world w_{16}:

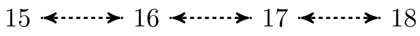

Figure 5.4: $M_{e,w_{16}}$, updating with e $(= p_{16} \vee p_{17} \vee p_{18})$ at world 16.

and at world w_{18}:

[23] Note that both standard DEL and standard Bayesian update lack this property. Such updates are insensitive to the world of evaluation, and hence transparent, according to the terminology of Chapter 2.

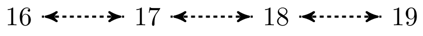

Figure 5.5: $M_{e,w_{18}}$, updating with e ($= p_{16} \vee p_{17} \vee p_{18}$) at world 18.

When we update with e at world w_{16}, we cannot eliminate the close world w_{15} from the updated model. This is why the model in Figure 5.4 contains world w_{15}. Similarly, world w_{19} remains in the updated model if we update with e at world w_{18} (in Figure 5.5). The key issue is rather trivial: inexact updates cannot eliminate close worlds, and the set of worlds that count as close changes according to the world of evaluation.

In the above models, relative to observation e, w_{17} is the good case while w_{16} and w_{18} are the bad cases (which are not, however, skeptical cases). In w_{17}, having the veridical experience of e results in knowing that e. To see why, note that in Figure 5.3, Ann knows e, as e is true everywhere in that model. Formally: $M, w_{17} \models [e]Ke$, since $M, w_{17} \models e$ and $M_{e,w_{17}}, w_{17} \models Ke$. In the not so good case, w_{16}, having the e experience does not result in knowing e, but the weaker $e \vee p_{15}$. Formally: $M, w_{16} \models [e]K(e \vee p_{15})$ (consult the updated model in Figure 5.4 to see why). Similarly, $M, w_{18} \models [e]K(e \vee p_{19})$. In the good case, Ann gets the most out of the evidence — in the other cases, she gets less.

Recall that the actual world in the example is indeed w_{17}. Thus, since $M, w_{17} \models [e]Ke$, the model satisfies what I have previously called the **main assumption**. Clearly, $\mathbf{P1}'$, $[e]K\neg p_{15}$ holds as well in w_{17}. Consider then the truth of $\mathbf{P2}'$: $M, w_{17} \models K[e](K\neg p_{15} \rightarrow \neg p_{16})$. The only world that can witness the falsity of the known conditional is w_{16}, in which the consequent is false. The question is then whether $M, w_{16} \models [e](K\neg p_{15} \rightarrow \neg p_{16})$ holds. The answer is yes, because the antecedent is false: at world w_{16}, having the e experience does not result in knowing $\neg p_{15}$ (consult Figure 5.4). Thus, $\mathbf{P2}'$ is true in my model; the agent knows the epistemic effects of her margin-for-error. Moreover, the KK principle holds in all the models of Figures 5.2 to 4.5, as in these models the epistemic R relation is universal.[24] Static introspection is not a cause for concern.

Finally, and most importantly, note that (DI) fails in my model. We have that $M, w_{17} \models [e]Ke$, so according to the latter principle we should also have $K[e]Ke$ at w_{17}. But that is false, since, from the perspective of the initial model, Ann considers w_{16} to be epistemically possible. In w_{16} however $\neg[e]Ke$, thus, $M, w_{17} \models \neg K[e]Ke$.

The above line of reasoning is (again) rather familiar within externalist epistemology: although Ann is actually in the good case (w_{17}), and although the observation of e results in knowledge in the good case, *for all Ann knows*, she is in the bad case w_{16}, and in the bad case, the same experience will not result in the same state of knowledge. Since Ann does not initially know whether she is in the good case or not, she cannot initially know how the e experience will affect her knowledge state. Ann does not have dynamic introspection.

[24] One can also construct updated models which are not S5 models, rather only S4. The 5 axioms does not play any crucial rule in my argument.

This completes the basic semantic picture of inexact observations. The model I offered (Figure 5.2) shows that it is possible to satisfy the KK principle and prior knowledge of margin-for-error, while falsifying dynamic introspection. However, the model does not assume that the agent has unrestricted knowledge of margin-for-error posterior to the observation. This is because I reject the claim that agents have unrestricted knowledge of their margin for error both prior and posterior to observations. I explain this in Section 5.5.

5.4 Evidential probabilities and modesty

My picture can be supplemented with evidential probabilities, in the style mentioned in the introduction of this chapter. Since all my models are **S5** models, the prior probability on a model M and the probability at a world collapse into one. The question is rather how to update the probability function from an initial model M to its updated version $M_{\varphi,w}$. One natural answer is straightforward: assuming that P is the probability function defined on the initial model M, let $P_{\varphi,w}$, the probability function over the updated model $M_{\varphi,w}$, be defined as $P_{\varphi,w}(\cdot) = P(\cdot|W')$ (recall that W' is the set of worlds after the update). The suggestion is to update the probability function via conditionalization with the same set with which we update the epistemic model.

Thus, if each world w_i in the initial model was assigned $\frac{1}{60}$, i.e. $P(w_i) = \frac{1}{60}$, the updated probability function $P_{e,w_{17}}$, corresponding to the model in Figure 5.3, will assign $\frac{1}{3}$ to each world. In the less than perfect worlds of w_{18} and w_{16} (Figures 5.4,5.5), each world will have $\frac{1}{4}$ weight. The point is that in the actual world, after looking, Ann has 100% confidence both in e and in Ke: there is not split between first order and higher-order evidence in my account of the unmarked clock example.

What about modesty? Doesn't Ann become immodest by having 100% confidence in her own evidence? No, because Ann's modesty is found in the fact that (DI) is false. What makes Ann's perceptual knowledge modest is the fact that she cannot predict in advance the knowledge state the observation will generate. Even though $\langle e \rangle Ke$ is true at world w_{17}, Ann is not able to predict, before looking at the clock, that having a veridical e experience will result in knowledge that e. In this respect, the unmarked clock example is analogous to the simpler **Tree** example. In both cases, the agents' epistemic modesty is a result of the externalist epistemology: cases in which knowledge is generated by a safe, reliable process although the agents do not know that the process is such. This dynamic type of modesty can be manifested in probabilistic terms, as I will now show.

Williamson explicates the *synchronic reflection principle* as the equation
Synchronic reflection: $P_w(X|\{u \in W : P_u(X) = c\}) = c$.
(see Williamson 2014, appendix). **Synchronic reflection** states that the probability of proposition X, given the proposition that the probability of X is c, is c. Williamson shows that on **S5** models, **Synchronic reflection** holds, but that on non-transitive or non-symmetric models, the principles

fails. Since Williamson's intended model (see Figure 5.1) is not transitive, Williamson concludes that inexact observations falsify synchronic reflection. Others, like Elga (2013) have formulated similar (synchronic) reflection principles and argued that cases of inexact observation falsify such principles.

Since the R relation in my models **S5** is an equivalence relation, the synchronic reflection principle is not violated on them. On the other hand, it is easy to see that my models falsify a version of the *diachronic reflection principle*. Diachronic reflection, in the context of dynamic epistemic logic, states that the probability of φ given the proposition that after updating with ψ, my probability in φ is c, is c. A standard representation of diachronic reflection states that:
$Pr_{t_0}(\varphi | Pr_{t_1}(\varphi) = c) = c$
where t_0 is the current time and t_1 is some future point in time. Within dynamic epistemic logic we make the assumption that time t_1 is the time after the update with some ψ. Expanding the way Williamson explicates synchronic reflection in static epistemic logic, in dynamic epistemic logic the diachronic reflection principle can be explicated as the following equation
Diachronic Reflection: $P_w(\varphi| \ \{u \in W | P_{\psi,u}(\varphi) = c\} \ ) = c.$

In words: the probability I assign to φ, given the proposition that after updating with ψ I assign c to the probability of φ, is c. The intuition behind the principle is that my future self (myself after further updates) is more informative than my current self. Thus, I should regard my future self as an expert, and align my current probabilities with my future probabilities.

The question then is what is the relation between the **Diachronic reflection** and the updates of dynamic epistemic logic. In part 2 of the appendix I prove that standard dynamic epistemic logic, i.e. a logic in which the updates are exact, the diachronic reflection is true, given the assumption that the agent has full static introspection. Diachronic reflection fails, however, in the case of inexact updates, even if the agent has full static introspection. Note for instance that in Figure 5.2, the set of worlds in which after inexactly updating with e assigns p_{17} probability $\frac{1}{3}$ consists only of w_{17}; $\{u \in W \mid P_{e,u}(p_{17}) = \frac{1}{3}\} = \{w_{17}\}$. But $P_w(p_{17}|\{w_{17}\}) = 1$, so $P_w(p_{17}|\{u \in W \mid P_{e,u}(p_{17}) = \frac{1}{3}\}) = 1$ which is a counterexample to diachronic reflection. The failure of diachronic reflection might not come as a big surprise, given that the future evidence in the example does not form a partition. It is however important to understand why and in what way the evidence is non-partitional. Moreover, we see that we can model the situation such that that body of evidence at any given state forms a partition. The phrase *non-paritional evidence* is quite ambiguous.[25]

The conclusion to draw from these formal considerations is that cases of inexact observations give rise to counterexamples to the diachronic reflection principle. One should not conclude from this that synchronic reflection is true. Rather, my point was to show that cases of inexact observations, when understood dynamically, are compatible with synchronic reflection but not with diachronic reflection. Thus, one cannot use such examples to reject synchronic reflection (as is done in Williamson 2014, Elga 2013). The probabilistic conclusion aligns with the logical conclusion from

[25] For more on this point, see also Weisberg (2007: 184) and Briggs (2009).

previous sections: in general, I argue, cases of inexact observation give rise to failures of dynamic principles (dynamic introspection, diachronic reflection); the corresponding static principles (the KK principle, synchronic reflection), whether true or false, are not threatened by such scenarios.

5.5 Back to Safety

I have offered an alternative explanation to the tension between inexact observations and introspection, based on the language of dynamic epistemic logic and my novel semantics of inexact updates. In particular, I reformulated Williamson's margin-for-error principle with inexact updates and argued that it fits naturally with the commitments of externalist theories of knowledge. The fact remains that Williamson's three static premises are inconsistent in (static) epistemic logic, so I must reject one of them. I reject **P2**, Williamson's formulation of the agent's knowledge of their own margin-for-error. In this section I explain how knowledge of margin-for-error is compatible with a rejection of **P2**. I argue that Williamson's reasoning pattern (presented in the derivation in Section 5.1) cannot be iterated. Even in cases where the agent knows her margin-for-error, this knowledge can be only used once. The objection I present here is in the same spirit of earlier critics of Williamson's argument, notably Sharon and Spectre (2008) and Dokic and Égré (2009), but my overall dynamic analysis allows to present this type of criticism in a novel, more comprehensive perspective.

Up to now, we have formulated knowledge of margin-for-error (of one unit) as $K[e](K\neg p_i \rightarrow \neg p_{i+1})$. This formula describes the prior *de-dicto* knowledge Ann has about the effect of making an observation e with a margin-for-error of 1 unit. This prior de-dicto knowledge does not imply that posterior to the observation event, Ann has the *de-re* knowledge that *that* event was an inexact observation event with a margin-for-error of 1 unit. After all, Ann can be uncertain as to her exact margin-for-error at that particular observation event. To avoid this complication, we can simplify things by considering a margin-for-error of an even smaller value. Consequently, we can assume that Ann can know that the observation event she experienced followed a margin-for-error of that small unit. So let us assume that Ann is in a position to know that, posterior to the observation event, her perceptual margin-for-error for that event must *at least* be 0.1. Thus, we assume that the formula $[e]K(K\neg p_i \rightarrow \neg p_{1+0.1})$ is true.[26]

According to Williamson, if the KK principle is true, then after the observation, Ann can use her margin-for-error knowledge again and again to rule out p_{16}, $p_{16.1}, p_{16.2}$... etc. I agree that Ann can use her margin-for-error once to rule out the p_{16} possibility, but Williamson is wrong, I argue, in assuming that this reasoning process can be iterated. After the observation *and* Ann's reasoning process Ann considers it possible that she knows that it is not p_{16} and that the world is actually 16.1. In other symbols, $\hat{K}(K\neg p_{16} \wedge p_{16.1})$ is the case. This is a counterexample to Williamson's **P2** for margin-for-error 0.1. The sentence $\hat{K}(K\neg p_{16} \wedge p_{16.1})$ says that Ann considers it possible that she is on

[26] Nothing hinges on this choice of values. For whatever value of margin-for-error we choose, we cannot iterate the Williamsonian reasoning pattern from section 1.1.

the 'edge', knowing $\neg p_{16}$ very close to p_{16} (in a world where $p_{16.1}$ is true). However, this conclusion should not count as violating the margin-for-error principle and the safety intuition behind it, since the margin-for-error principle is meant to describe the agent's knowledge state *resulting from* an inexact observation, *not* her knowledge state in general. Recall that an inexact observation is safe iff as a result of the observation the actual world is surrounded by a large enough buffer zone of close possible worlds. In this sense, Ann's knowledge gained directly by the inexact observation of e is safe. This knowledge state is then combined with Ann's background knowledge about her own margin-for-error, resulting in a new knowledge state in which the 0.1 subintervals on the sides have been ruled out. Now, this new knowledge state should not be constrained by observational inexactness anymore, as it is not purely observational knowledge at that point. The margin-for-error principle only applies to the knowledge state obtained as a result of an inexact observation, it does *not* apply to knowledge states obtained by observation *together* with other non-observational means. The margin-for-error principle is a principle about *perceptual* knowledge; it is not a principle about knowledge *in general*.

To emphasize this point, consider the following variation of the unmarked clock example. Suppose that before looking at the clock, Ann does not form any beliefs about her margin-for-error. She looks at the clock and comes to know that the minute hand is somewhere in the interval 16-18; suppose it actually points to 17 and that Ann's actual margin-for-error is 1. Now Ann's optometrist shows up and tells her that given the conditions of *the observation she just made*, she cannot reliably perceptually distinguish 0.1 distance on the clock face: it is the case that if the minute hand points to p_i, then for all Ann knows it points to $p_{i\pm 0.1}$. In other words, the optometrist tells Ann that her margin-for-error is *at least* 0.1. Since Ann's optometrist is a known reliable source (let's assume so), Ann comes to know that her margin-for-error is at least 0.1. Ann then uses her knowledge gained from the optometrist (and the KK principle) to cut her uncertainty interval by 0.1 on both sides, coming to know that the minute hand is somewhere in the interval 16.1-17.9. More specifically, we assume that Ann knows that the clock is not pointing to 15.9 ($K\neg p_{15.9}$) as a result of the inexact observation, and that she knows that if she knows that it does not point to 15.9 it cannot be pointing to 16 ($K(K\neg p_{15.9} \to \neg p_{16})$), by the optometrist testimony. By the KK principle, $KK\neg p_{15.9}$ obtains, and so by the assumed closure of knowledge, it follows that $K\neg p_{16}$. Likewise, from the assumptions that $K\neg p_{18.1}$ (due to the inexact observation) and $K(K\neg p_{18.1} \to \neg p_{18})$ (due to the testimony), it follows that $K\neg p_{18}$. The remaining epistemic possibilities range from 16.1 to 17.9.

Importantly, note that Ann cannot reuse the knowledge she gained from her optometrist to conclude anything stronger; she cannot iterate the process. The optometrist did not convey the information that $p_{i+0.1} \to \hat{K}p_i$ is true in general; they conveyed the information that *after the observation Ann just made* it is true that $p_{i+0.1} \to \hat{K}p_i$. Once Ann uses the information she got from the optometrist, the sentence $p_{i+0.1} \to \hat{K}p_i$ becomes false, because the context has changed: now the K operator does not refer solely to knowledge after the observation, but to knowledge after making an observation and learning from testimony. The optometrist did *not* say "after anything you learn, it must be the case that $p_{i+0.1} \rightarrow \hat{K}p_i$", they said "after you make an inexact observation, it must be the case that $p_{i+0.1} \rightarrow \hat{K}p_i$." The latter statement is a true description of the effect of inexact observation. The former statement is a false and ungrounded description of Ann's general knowledge structure. Surely it is possible for Ann to come to know $K\neg p_i \wedge p_{i+0.1}$ by some other non-observational means.

Williamson thinks that there is something problematic about a situation in which Ann considers it possible that the clock points to 16.1 and at the same time she knows that is does not point to 16. Williamson believes that such a situation is in a direct conflict with the safety condition of knowledge. I disagree. Given the fine-grained dynamic analysis I propose, this situation can be explained. Consider the optometrist version of our story again. Ann starts by making an inexact observation and learning that the minute hand is between 16 and 18. After the optometrist informs Ann about her perceptual margin-for-error, she concludes that the minute hand must be between 16.1 and 17.9. When asked whether she thinks it is possible that the minute hand is in fact pointing to 16.1, she can respond: "for all I know, the minute hand is actually pointing to 16.1. I can tell that it is not pointing to 16 because this contradicts the observation I made together with the optometrist's information. But I cannot conclude anything stronger. In particular, it is possible, as far as I can tell, that I have initially observed that the minute hand is pointing somewhere between 16 to 18 and that it was actually pointing to 16.1. This state of affairs does not contradict my assumption that the optometrist spoke truly (i.e. the margin-for-error principle is correct). If I knew that my perceptual margin-for-error is larger than 0.1, I could have ruled out the possibility that the clock is actually pointing to 16.1. But I don't know that." In this context, I think there is nothing problematic about Ann's response. Ann's perceptual knowledge, obtained by an inexact observation, is safe. Ann's resulting knowledge state, after taking into account the optometrist's testimony, does not violate the safety requirement of perceptual knowledge (i.e. the margin-for-error principle), because it is not purely a perceptual knowledge state anymore.

Figure 5.6 graphically summarizes my analysis of the situation, incorporating the lessons from this section and section 5.3. It contains three (simplified) epistemic models and two updates: the top model represents Ann's initial epistemic state, before making any observation.[27] The middle model represents Ann's epistemic state after inexactly observing e but before learning about her margin-for-error. The transition between the top model and the middle model was explained in detail in section 5.3. The bottom model represents Ann's knowledge state after learning from the optometrist about her margin-for-error. It is obtained by eliminating all the possible worlds in the middle model in which $K\neg p_i \rightarrow \neg p_{i\pm 0.1}$ is false (i.e. by an exact update with $K\neg p_i \rightarrow \neg p_{i\pm 0.1}$). Note that it is only worlds w_{16} and w_{18} which are eliminated. In w_{16}, for instance, we have $K\neg p_{15.9} \wedge p_{16}$, contradicting the information the optometrist conveyed, so it is eliminated. The important thing to note is that

[27] The models are simplified because the P relation is not drawn.

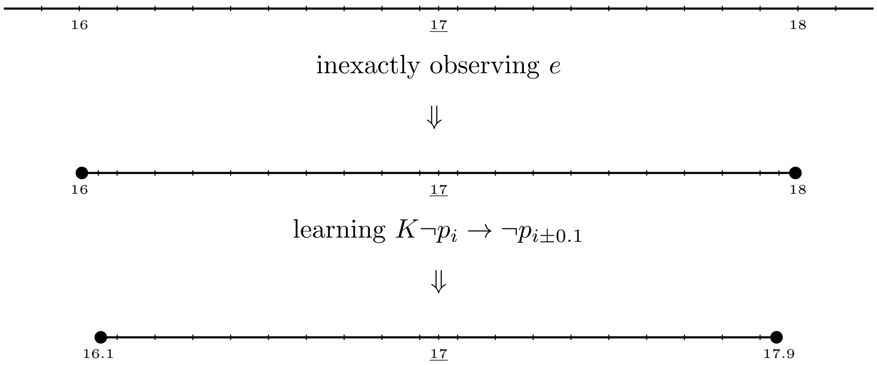

Figure 5.6: Two model transitions: Ann first inexactly observes e, then incorporates her margin-for-error knowledge.

when the optometrist announces that $K\neg p_i \to \neg p_{i\pm 0.1}$, they refer to the knowledge state in the middle model, the knowledge resulting from an inexact observation, not to any other knowledge state.[28] In world w_{17} of the bottom model, it is true that $\hat{K}(K\neg p_{16} \wedge p_{16.1})$. As I explained in the previous paragraph, this is not in conflict with the safety condition for perceptual knowledge. In the bottom model, the K is interpreted as the knowledge state obtained from a combination of an inexact observation and the information received from the optometrist. Therefore, the safety condition for perceptual knowledge (i.e. the margin-for-error principle) does not apply to this state.

The addition of the optometrist is not essential to my account. The only difference between the two versions of the story is that in Williamson's story, knowledge of margin-for-error is obtained before the observation (by reflection); in the optometrist story, this knowledge is obtained after the observation (by testimony). The stage (and method) in which the agent learns their margin-for-error should not affect their final knowledge state. I believe, however, that my modified version of the story makes it easier to recognize that the margin-for-error principle is true only for the knowledge state obtained by the inexact observation. Figure 6 can equivalently be interpreted as representing the process where Ann first observes e and then incorporates her background knowledge about her margin-for-error. Under my account, Ann can have knowledge of margin-for-error posterior to the observation. She can then use that knowledge once, but not more than that. Figure 5.6 represents this two step process of first making an inexact observation, and then incorporating one's knowledge of margin-for-error.

One might offer the following objection to my analysis: since Ann's final knowledge state is the result of both her inexact observation and her knowledge of her margin-for-error, her final knowledge state is inexact as well. And if her final knowledge state is inexact, then it must follow a margin-for-error principle by itself. If Ann knows this margin-for-error (and there is no reason to assume

[28] In the terminology of dynamic epistemic logic, the announcement made by the optometrist is not *successful*, because it is not known after the announcement (see Baltag and Renne 2016). This is an indication that the content of the announcement is context sensitive, in the sense used within dynamic epistemic logic (see, e.g. Holliday 2018).

otherwise), she can apply it again, come to know a stronger proposition, repeat the same reasoning again, and so potentially reach Williamson's contradictory conclusion.

There are several problems with the objection. First, the objection wrongly assumes that any knowledge state that is partially the result of an inexact observation must itself be an inexact knowledge state, and so follow a margin-for-error. Here is a counterexample to this assumption. Consider the following scenario: Ann makes an inexact observation and comes to know that the minute hand is not pointing to zero, $\neg p_0$. This is an assumption that both Williamson's story and my story can accommodate. Later, an Oracle tells Ann that the clock is not pointing in the range $p_1 - p_{16}$, nor does it point in the range $p_{18} - p_{59}$ (leaving only p_0 and p_{17} possible). After receiving the Oracle's information, Ann comes to know p_{17}, using her knowledge from the inexact observation, the Oracle's information, and her deductive abilities. Thus, Ann knows exactly where the minute hand is pointing to, and this knowledge state is the result of both the inexact observation and the Oracle's information. According to the assumption in the objection, Ann's final knowledge state is inexact, because it is partially the result of an inexact observation. If so, then her final knowledge state must follow a margin-for-error. But Ann's final knowledge seems to be exact in this scenario (she knows exactly where the minute hand is pointing), and so her knowledge does not follow a margin-for-error anymore. Thus, the scenario offers a counterexample to the claim that every knowledge state partially obtained by an inexact observation must be an inexact knowledge state that follows a margin-for-error. It is possible to reach a knowledge state that does not follow a margin-for-error from a previous knowledge state that does.

Second, even if we assume that the agent's final knowledge state is inexact, it is unclear how this leads to a contradiction. For the sake of the argument, grant the objector the assumption that—unlike the scenario described in the last paragraph—every method the agent has for obtaining knowledge is in some way inexact, and so follows some margin-for-error. Even with this assumption, it is far from clear that Williamson's neat contradictory derivation follows. The force of Williamson's argument comes from the fact that the perceptual margin-for-error principle $p_i \to \hat{K}p_{i+\epsilon}$ is so intuitive: clearly, there must be some ϵ such that I cannot visually discriminate between the minute hand being in position i (p_i) and position $i+\epsilon$ ($p_{i+\epsilon}$). Other margin-for-error principles (resulting from the agent's non-visual inexact methods of gaining knowledge) will not be so easy to accept, or even to formulate. Assume that Ann's deductive abilities are inexact, and so the knowledge Ann obtains by deduction follows some margin-for-error. To articulate such margin-for-error, one would first have to come up with a notion of similar possibilities for the outcomes of Ann's deductive inferences, and then formulate a margin-for-error principle based on that notion of similarity. There is no reason to assume that the latter notion of similarity will have anything to do with a notion of similarity based on metric distance (which we use for perceptual inexactness). The same can be said of other potential inexact methods of gaining knowledge, like inexact memory or inexact (i.e.

potentially unsafe) testimony. Even if Ann's knowledge state is governed by a further margin-for-error principle(s), there is no reason to assume that those margin-for-error have the form $p_i \rightarrow \hat{K}p_{i+\epsilon}$, which is crucial for Williamson's derivation of contradiction. It is the burden of the objector to offer a compelling story as to why a different margin-for-error principle resulting from another inexact method of gaining knowledge leads to a contradiction. And my response to such attempt will be similar to my earlier response: a margin-for-error principle is applicable for a particular epistemic state; once the margin for error is used by the agent, the epistemic state has changed, and there is no reason to assume that the same margin-for-error applies in the new state as well (a different margin-for-error might apply to the new situation, but there is no contradiction in that).[29]

Williamson's static formulation of knowledge of margin-for-error, $K(K\neg p_i \rightarrow \neg p_{i+\epsilon})$, is inadequate exactly because it does not capture the idea that the inner K in it refers to knowledge *after an inexact observation.* By reusing the margin-for-error principle in his derivation (presented in Section 5.1.1), Williamson implicitly assumes that the inner K in the principle describes a general knowledge state of the agent. As the optometrist story meant to convey, this is a mistaken assumption. My dynamic formulation of inexact perceptual knowledge fares better. Syntactically, the ability to scope the K in the inexact observation operator $[e]$ allows us to represent the knowledge which results from an inexact observation. Semantically, the mechanics of my inexact updates allow me to formally connect the margin-for-error principle with the epistemic *result* of an inexact observation. As a consequence, we get a broader picture of *inexact observational knowledge*, as the knowledge that results from inexact observations. I conclude that my dynamic story does a better job in capturing the epistemic aspects of inexact observations.

As I mentioned earlier, the objection that the margin-for-error principle is context specific and should not be assumed in an unrestricted form already appears in the literature (Sharon and Spectre 2008, Dokic and Égré 2009). For this reason, I would like to stress the main differences between my approach and earlier criticisms. First, unlike Dokic and Égré (2009) my analysis does *not* rely on distinguishing between two types of knowledge operators, perceptual and reflective (see also Halpern 2008, Égré and Bonnay 2008, Sharon and Spectre 2008 for similar proposals). My conceptual, syntactic and semantic treatment of the knowledge operator is uniform. My dynamic analysis, however, does allow me to distinguish between the behavior of different sources of information, and to associate different safety requirements with different sources. Moreover, and unlike earlier criticisms, my formal framework cannot be accused of being *ad-hoc*, given that dynamic epistemic logic is an independently motivated formal framework.[30]

Second, and more importantly, unlike earlier criticisms, my analysis manages to capture the

[29] My response essentially appeals to a quantifier shift fallacy. To get a contradictory derivation, the objector needs to assume that there is one type of margin-for-error principle for every way of gaining inexact knowledge. I argue that for every way of gaining inexact knowledge, there is some type of margin-for-error principle. Since I see no reason to assume that the different margin-for-errors have the same structure, I don't see how one type of margin-for-error can be repeatedly applied to obtain a contradiction.

[30] See Dokic and Égré (2009:19) for a response to the *ad-hoc* accusation.

compelling elements of Williamson's story. Like Williamson, I believe that there is a tension between inexact observations and introspection, and that this tension is worthy of a philosophical explanation. Furthermore, I believe that Williamson's argument is persuasive *because* it offers an explanation to this tension. Earlier objections have pointed out the flaws in assuming that the margin-for-error principle holds unrestrictedly, but they have not offered an alternative positive account to explain in what ways inexactness is incompatible with introspection. The dynamic account I present in this paper shows how inexact observations are incompatible with a dynamic form of introspection. The account also retains the buffer zone intuition that plays such an important role in Williamson's framework. However, as the last few sections have shown, one can accommodate and explain the tension between inexactness and introspection without the need to reject KK.

5.6 Concluding remarks

Inexact observations are important when it comes to introspection; they are important for dynamic introspection. In sections 5.2 and 5.3 I have shown how inexact observations are in conflict with dynamic introspection. In section 5.4 I showed the tension between diachronic reflection and inexact observations. In section 5.5 I have further argued that Williamson reaches a wrong conclusion by suppressing the dynamic aspects of inexact observations. My argument extends to any account that takes epistemic indistinguishability to be non-transitive in situations of inexact observations,[31] and more broadly to any account that, following Williamson, assumes that inexact observations create a synchronic conflict between first- and higher-order evidence. Within his static formulation, Williamson motivates **P2** with the idea that knowledge requires safety, and safety requires an error free buffer zone; knowledge can never be obtained at a world 'on the edge.' My account of inexact observations fully adheres to the buffer zone intuition that motivates Williamson. Every inexact observation event leads to an updated model in which the actual world is surrounded by close but epistemically possible worlds that act as a buffer zone (consult the position of the actual world in Figures 5.3 to 5.5). No observation event can put the actual world at the 'edge' of the updated model – and thereby the agent in an epistemically dangerous place. But my account also shows that one can follow the safe buffer zone intuition without accepting a strong premise like **P2**. If one wants to stick with the truth of **P2**, one cannot just cite margin-for-error type safety considerations. Even in situations where the agent is able to use their margin-for-error knowledge to conclude something stronger about what they know, such reasoning cannot be iterated. This is because the margin-for-error principle only describes the epistemic effect of inexact observations; it does not describe the effects of inexact observations after they are combined with non-observational knowledge. In such cases, knowledge of margin-for-error cannot be used to come to know anything stronger.

A more general conclusion is that in the context of an externalist epistemology, reasoning about

[31] E.g. Elga 2013, Salow and Ahmed 2017.

epistemic updates requires special care. In standard dynamic epistemic logic, update operators are *transparent* to the agent, in the sense that the behavior of the update is the same inside and outside the scope of knowledge.[32] This idealized assumption is exemplified in two DEL axioms, known as *no-miracles*, $\langle\varphi\rangle K\psi \to K[\varphi]\psi$ and *perfect recall*, $K[\varphi]\psi \to [\varphi]K\psi$ (see Chapter 2). Note how the two axioms allow to switch the order of the knowledge and update modalities. In my account of *inexact updates*, these two axioms fail. This is important, as it suggests that within externalism, reasoning about epistemic updates is not transparent, but *opaque*. This should make sense: updates behave differently at the good and bad case, and when we don't know whether we are in the good case, there is no reason to assume that the update will behave in the ways we expect. We don't always know how new evidence will affect us, nor do we always know what evidence brought us to our current state. This *opacity* is crucial for understanding externalist theories of knowledge and some of the puzzles associated with them.

Appendix 1: Deriving an absurdity in the unmarked clock example with dynamic introspection

Recall that e abbreviates $p_{16} \vee p_{17} \vee p_{18}$ and nothing else. We assume the principles:

$[e]Ke$	**main assumption**
$K[e](K\neg p_{15} \to \neg p_{16})$	**P2$'$**
$[\varphi]K\varphi \to K[\varphi]K\varphi$	dynamic introspection
$\langle\varphi\rangle\psi \leftrightarrow [\varphi]\psi \wedge \varphi$	**Partial function**

We further assume that both K and $[\varphi]$ obey:

$\vdash \varphi \Rightarrow \vdash \Box\varphi$	nec. rule
$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	$\to$-distribution
$\Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$	$\wedge$-distribution

First we show that update-knowledge closure holds:

Lemma 1.1: $\vdash K[\varphi]K(\alpha \to \beta) \wedge K[\varphi]K\alpha \to K[\varphi]K\beta$

$\vdash K(\alpha \to \beta) \wedge K\alpha \to K\beta$	theorem of epistemic logic
$\vdash [\varphi](K(\alpha \to \beta) \wedge K\alpha \to K\beta)$	nec. rule of $[\varphi]$
$\vdash [\varphi](K(\alpha \to \beta) \wedge K\alpha) \to [\varphi]K\beta$	distribution of $[\varphi]$
$\vdash [\varphi]K(\alpha \to \beta) \wedge [\varphi]K\alpha \to [\varphi]K\beta$	distribution of $[\varphi]$
$\vdash K([\varphi]K(\alpha \to \beta) \wedge [\varphi]K\alpha \to [\varphi]K\beta)$	nec. rule of K
$\vdash K[\varphi]K(\alpha \to \beta) \wedge K[\varphi]K\alpha \to K[\varphi]K\beta)$	distribution of K

[32] A similar point holds for Bayesian epistemology, in which it is assumed that the posterior epistemic state is transparent to the agent prior to the update (as a prior conditional state).

Now, we show how to derive the problematic conclusion $K[e]\neg p_{16}$ (in line 10) from our assumptions. Line 6. establishes what I called earlier the auxiliary assumption.:

1. $[e]Ke$ **main assumption**
2. $K[e]Ke$ by (DI)
3. $e \to \neg p_{15}$ analytic truth (or model validity)
4. $K(e \to \neg p_{15})$ nec. rule of K
5. $[e]K(e \to \neg p_{15})$ nec. rule of $[\varphi]$
6. $K[e]K(e \to \neg p_{15})$ nec. rule of K
7. $K[e]K\neg p_{15}$ by Lemma 1.1
8. $K[e](K\neg p_{15} \to \neg p_{16})$ assumption
9. $K([e]K\neg p_{15} \to [e]\neg p_{16})$ →-distribution for [ ]
10. $K[e]\neg p_{16}$ →-distribution for K

Appendix 2. Diachronic Reflection and Dynamic epistemic logic

Given a Kripke model M, an exact φ-updated model is defined as $M_\varphi = (W_\varphi, R_\varphi, V_\varphi)$, where $W_\varphi = W \cap \{u \in W : M, u \models \varphi\}$ and the rest is restricted accordingly (equivalently, we assume that P is empty). This is just the familiar public announcement (PA) update from dynamic epistemic logic. The resulting logical system is known as PAL (public announcement logic). See Baltag & Renne (2016).

Note that this notion of updating is factive: you only change the model if φ is true.

Further note that PAL model transformations are partial functions. Given an M, there is at most 1 updated M_φ but not more.

We define the updated P_φ as $P_\varphi(\cdot) = P(\cdot|W_\varphi)$.

We write $P_{\varphi,w}$ for the probability function at world w after updating with φ. It is defined as:
$P_{\varphi,w}(\cdot) = P_\varphi(\cdot|R_\varphi(w))$.
Like in the static case. Note that if $w \notin W_\varphi$, then $R_\varphi(w) = \emptyset$. In that case, the probability $P_{\varphi,w}(\cdot)$ is undefined. In the contrapositive: if $P_{\varphi,w}$ is defined, then $w \in W_\varphi$. This observation will be used below.

Recall that we explicate the **diachronic reflection principle** as
DRP: $P_w(X|\{u \in W : P_{\varphi,u}(X) = c\}) = c$
Given the proposition that after updating with φ your probability in X is c, your current probability of X is c.

Example: Given the information that after listening to a (truth teller) weather forecaster (φ) your probability of rain tomorrow (X) is .9 (c), your current probability that it will rain tomorrow is .9.

Fact: DRP holds on S5-PAL models. Introspective PAL agents have **DRP**.

We wish to show:
$P_w(X|\{u \in W : P_{\varphi,u}(X) = c\}) = c$.
Note that this expression is only defined if $\{u \in W : P_{\varphi,u}(X) = c\} \neq \emptyset$. So we assume that throughout.

We now prove that
(i): $\{u \in W : P_{\varphi,u}(X) = c\} = W_\varphi$.
$\Rightarrow$: Assume $v \in \{u \in W : P_{\varphi,u}(X) = c\}$. So $P_{\varphi,v}(X) = c$. So $P_{\varphi,v}$ is defined, thus $v \in W_\varphi$ (see earlier observation).
$\Leftarrow$: Since we assume that $\{u \in W : P_{\varphi,u}(X) = c\} \neq \emptyset$, there is some u s.t. $P_{\varphi,u}(X) = c$. Since the M_φ model is S5, all other worlds $v \in W_\varphi$ are also s.t. $P_{\varphi,v}(X) = c$. So assume that $w \in W_\varphi$. It follows that $P_{\varphi,w}(X) = c$. So $w \in \{u \in W : P_{\varphi,u}(X) = c\}$.

Now:
$P_w(X|\{u \in W : P_{\varphi,u}(X) = c\}) = P_w(X|W_\varphi)$ by (i), and
$P_w(X|W_\varphi) = P_{\varphi,w}(X)$ by definition, so
(ii): $P_w(X|\{u \in W : P_{\varphi,u}(X) = c\}) = P_{\varphi,w}(X)$.

If $P_{\varphi,w}(X)$ is defined, then
$w \in W_\varphi$ (given the earlier observation), and
$\{u \in W : P_{\varphi,u}(X) = c\} = W_\varphi$ (by (i)), so
(iii): $P_{\varphi,w}(X) = c$.

(ii) and (iii) imply $P_w(X|\{u \in W : P_{\varphi,u}(X) = c\}) = c$.

Bibliography

— Arntzenius, Frank (2003). Some Problems for Conditionalization and Reflection. *Journal of Philosophy.* 100 (7):356-370.

— Aucher G., Herzig A. (2010). Exploring the power of converse events. in *Dynamic Epistemology: Contemporary Perspectives*, Springer, Synthese Library.

— Aumann, R. Interactive epistemology I: Knowledge. (1999). *Game Theory* 28, 263–300

— Baltag, Alexandru (2016). To Know is to Know the Value of Variable. In Lev Beklemishev, Stéphane Demri & András Máté (eds.), *Advances in Modal Logic*, Volume 11. CSLI Publications. pp. 135-155.

— Baltag, Alexandru and van-Benthem, Johan. (2018). Some thoughts on the logic of imprecise observations. *Tsinghua Philosophy Journal*, Tsinghua University, Beijing..

— Baltag, A., Bezhanishvili, N., Özgün, A. Smets, S. (2019) A Topological Approach to Full Belief. Journal of Philosophical Logic 48, 205–244.

— Baltag, A., Gierasimczuk, N. & Smets, S., (2015). On the Solvability of Inductive Problems: A Study in Epistemic Topology. *Proceedings of the 15th Conference on Theoretical Aspects of* Rationality and Knowledge (TARK 2015), R. Ramanujam (ed.), Chennai: Institute of Mathematical Sciences, pp. 65–74

— Baltag, Alexandru and Renne, Bryan. (2016) Dynamic Epistemic Logic. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition).

— Baltag, Alexandru and Smets, Sonja. (2008) A qualitative theory of dynamic interactive belief revision, in G. Bonanno, W. van der Hoek, and M. Wooldridge (eds.), *TLG 3: Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Volume 3 of Texts in logic and games, pp. 11–58, Amsterdam: Amsterdam University Press.

— Ballarin, Roberta. (2021) Modern Origins of Modal Logic. in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.).

— van Benthem, Johan (2003). Conditional probability meets update logic. *Journal of Logic, Language and Information.* 12 (4):409-421.

— van Benthem, Johan (2010). *Modal Logic for Open Minds.* Center for the Study of Language and Information (CSLI). Stanford.

— van Benthem, Johan (2012). *Logical Dynamics of Information and Interaction.* Cambridge University Press.

— van Benthem, Johan (2014). *Logic in Games.* Cambridge, MA: The MIT Press.

— van Benthem, Johan. Fernández-Duque, David & Pacuit, Eric. (2014) Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic*, Volume 165, Issue 1,Pages 106-133, ISSN 0168-0072.

— van Benthem, Johan; Gerbrandy, Jelle; Hoshi, Tomohiro & Pacuit, Eric (2009). Merging frameworks for interaction. *Journal of Philosophical Logic* 38 (5):491-526.

— van Benthem, Johan and Klein, Dominik. Logics for Analyzing Games., *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.).

— van Benthem, J., J. van Eijck, and B. Kooi. (2006) Logics of communication and change, *Information and Computation,* 204(11): 1620–1662.

— van Benthem, J., & Bezhanishvili, G. (2007). Modal logics of space. In *Handbook of spatial logics* (pp. 217–298). Springer — Berto, Francesco and Hawke Peter. Knowability Relative to Information, *Mind*, Volume 130, Issue 517, January 2021, Pages 1–33.

— Bird, A. (2018), Evidence and Inference. *Philosophy and Phenomenological Research*, 96: 299-317.

— Bird, Alexander & Pettigrew, Richard (2019). Internalism, Externalism, and the KK Principle. *Erkenntnis*:1-20.

— Bjorndahl Adam (2018). The Epistemology of Nondeterminism. In: Moss L., de Queiroz R., Martinez M. (eds) *Logic, Language, Information, and Computation.* WoLLIC 2018. Lecture Notes in Computer Science, vol 10944. Springer, Berlin, Heidelberg.

— Bjorndahl, Adam (2018). Topological Subset Space Models for Public Announcements. In Hans van Ditmarsch & Gabriel Sandu (eds.), *Jaakko Hintikka on Knowledge and Game Theoretical Semantics.* Springer. pp. 165-186.

— Bjorndahl, Adam (2020). Knowledge Second. *Res Philosophica* 97 (4):471-487.

— Blackburn, Patrick ; de Rijke, Maarten & Venema, Yde (2002). Modal Logic. Cambridge University Press.

— Bonnay, Denis and Egre, Paul (2008). Margins for Error in Context. in M. Garcia- Carpintero & M. Kölbel (eds.), *Relative Truth*, pp. 103-127, Oxford University Press.

— Bonnay, Denis and Egre, Paul (2009). Inexact Knowledge with Introspection. *Journal of Philosophical Logic.* 38 (2), pp. 179-228.

— Bonnay, Denis and Egre, Paul (2011). Knowing one's limits - An analysis in Centered Dynamic Epistemic Logic. in P. Girard, M. Marion and O. Roy (eds), *Dynamic Formal Epistemology*, Springer, pp. 103-126

— Boh, Ivan (1993). *Epistemic Logic in the Later Middle Ages.* Routledge.

— Bradie, Michael and William Harms (2020). Evolutionary Epistemology. *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.).

— Briggs, R.A. (2009). Distorted reflection. *Philosophical Review* 118 (1):59-85.

— Burge, Tyler, (1979) Individualism and the Mental, *Midwest Studies in Philosophy*, 4: 73–121. doi:10.1111/j.1475-4975.1979.tb00374.x

— Brueckner, A., and Fiocco, M.O. (2002). Williamson's Anti-Luminosity Argument. *Philosophical Studies* 110: 285-293.

— Byrne, Alex and Logue, Heather. (2008). Either/Or. in *Disjunctivism: Perception, Action, Knowledge.* Adrian Haddock and Fiona Macpherson (eds.), Oxford: Oxford University Press, pp. 57–94.

— Carter, A.J., Gordon. C.E and Jarvis B. (eds), (2017). *Knowledge First: Approaches in Epistemology and Mind.* Oxford University Press.

— Chagrov, A., & Zakharyashchev 1992,, so the question of the M. (1992). Modal Companions of Intermediate Propositional Logics. Studia Logica: An International Journal for Symbolic Logic, 51(1), 49-82.

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

— Christensen, David (2010b). Rational Reflection. *Philosophical Perspectives* 24 (1):121-140.

— van Cleve, James, 2003. Is Knowledge Easy—or Impossible? Externalism as the Only Alternative to Skepticism", in Steven Luper (ed.). *The Skeptics: Contemporary Essays.* Aldershot: Ashgate, pp. 45–59.

— Cohen, Michael (2020). The problem of perception and the no-miracles principle. *Synthese.* https://doi.org/10.1007/s11229-020-02772-3

— Cohen, Michael. (2021a) Opaque Updates. *Journal of Philosophical Logic* 50, 447–470 .

— Cohen, Michael. (2021b) Inexact knowledge and dynamic introspection. Synthese.
https://doi.org/10.1007/s11229-021-03033-7

— Cohen, Michael. Wang, Yanjing and Wen Tang. (2021) De-re updates. In *Proceedings TARK 2021.* DOI: 10.4204/EPTCS.335.9

— Cohen, Stewart (2003). Basic Knowledge and the Problem of Easy Knowledge. *Philosophy and Phenomenological Research*, 65(2): 309–329.

— Cohnitz, Daniel & Haukioja, Jussi (2013). Meta-Externalism vs Meta-Internalism in the Study of Reference. *Australasian Journal of Philosophy* 91 (3):475-500.

— Comesaña, J., and Kantin, H. (2010). Is evidence knowledge? *Philosophy and Phenomenological Research*, 80, 447–454.

— Conee, E. (2005). The comforts of home. *Philosophy and Phenomenological Research* 70: 444-451.

— Das, Nilanjan & Salow, Bernhard (2018). Transparency and the KK Principle. *Noûs* 52 (1):3-23.

— Demey, L. (2015). The dynamics of surprise, *Logique et Analyse* 58, 251-277.

— Descartes, René, *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, John Cottingham, Robert Stoothoff, and Dugald Murdoch (trans.), vol. 2, Cambridge: Cambridge University Press, 1985

— van Ditmarsch, H.P. *Knowledge games.* Ph.D. thesis, University of Groningen, 2000. ILLC Dissertation Series DS-2000-06.

— Dorst, Kevin (2019). Abominable KK Failures. *Mind.* 128 (512):1227-1259.

— Dorst, Kevin (2020). Evidence: A Guide for the Uncertain. Philosophy and Phenomenological Research 100 (3):586-632.

— Dretske, Fred (2005). Is Knowledge Closed Under Known Entailment? The Case Against Closure. In Matthias Steup & Ernest Sosa (eds.), *Contemporary Debates in Epistemology.* Blackwell. pp. 13-26.

— Dutant, Julien (2007). Inexact Knowledge, Margin for Error and Positive Introspection. *Proceedings of Tark XI.*

— van Eijck J., Gattinger M., Wang Y. (2017) Knowing Values and Public Inspection. In: Ghosh S., Prasad S. (eds) *Logic and Its Applications. ICLA* 2017. Lecture Notes in Computer Science, vol 10119. Springer, Berlin, Heidelberg.

— Elga, Adam (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies* 164 (1):127-139.

— Fagin, Ronald ; Y. Halpern, Joseph ; Moses, Yoram Vardi, Moshe (1995). Reasoning About Knowledge. MIT Press.

— Feldman, Richard, and Earl Conee. Internalism Defended. (2001). *American Philosophical Quarterly* 38.1 : 1–18.

— van Fraassen, Bas. (1984). Belief and the Will. *Journal of Philosophy* 81:235–256.

— Fratantonio, G., and McGlynn, A. (2018). Reassessing the Case against Evidential Externalism. In V. Mitova (Ed.), *The Factive Turn in Epistemology* (pp. 84-101). Cambridge: Cambridge University Press.

— Girard, P., F. Liu, and J. Seligman (2012). General dynamic dynamic logic. In T. Bolander, T. Brauner, S. Ghilardi, and L. Moss (Eds.), *Proceedings of the 9th International Conference on Advances in Modal Logic (AiML'12)*, pp. 239–260. London: College Publications.

— Goldman, Alvin I. What Is Justified Belief?. (1979) in George S. Pappas (ed.), *Justification and Knowledge: New Studies in Epistemology*, Dordrecht: Reidel, pp. 1–25.

— Goldman, Alvin. (2009). Williamson on knowledge and evidence. In Patrick Greenough, Duncan Pritchard Timothy Williamson (eds.), *Williamson on Knowledge.* Oxford University Press. pp. 73-91.

— Goldman, Alvin and Beddor, Bob, (2016) Reliabilist Epistemology. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition).

— Goodman, J. and Salow, B. (2018). Taking a chance on kk. *Philosophical Studies*, 175(1):183–196.

— Greco, John (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity.* Cambridge University Press.

— Greco, Daniel (2014a). Could KK Be OK? *Journal of Philosophy* 111 (4):169-197.

— Greco, Daniel (2014b). Iteration and Fragmentation. *Philosophy and Phenomenological Research.* 88 (1):656-673.

— Greco, Daniel (2015a). Iteration Principles in Epistemology II: Arguments Against. Philosophy Compass 10 (11):765-771.

— Greco, Daniel (2015b). Iteration Principles in Epistemology I: Arguments For. *Philosophy Compass* 10 (11):754-764.

— Greco, Daniel (2016). Safety, Explanation, Iteration. *Philosophical Issues* 26 (1):187- 208.

— Greco, John (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity.* Cambridge University Press.

— Halpern, j., Samet, d., and Segev, e. (2009). Defining knowledge in terms of belief: the modal logic perspective. The review of symbolic logic, 2(3), 469-487.

— Hansen, N., Porter, J., & Francis, K. (2021). A Corpus Study of "Know": On The Verification of Philosophers' Frequency Claims about Language. *Episteme,* 18(2), 242-268.

— Hasan, Ali and Fumerton, Richard (2016) Foundationalist Theories of Epistemic Justification. *The Stanford Encyclopedia of Philosophy.*

— Hawthorne, John & Magidor, Ofra (2009). Assertion, Context, and Epistemic Access -ibility. *Mind.* 118 (470):377-397.

— Hedden, Brian (2015). Time-Slice Rationality. *Mind* 124 (494):449-491.

— Hintikka, Jaakko (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Ithaca: Cornell University Press.

— Holguín, Ben (2019). Indicative Conditionals and Iterative Epistemology. Manuscript.

— Holliday, Wesley (2018). Knowledge, Time, and Paradox: Introducing Sequential Epistemic Logic. In *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics.* Springer Verlag.

— Horowitz, Sophie (2014). Epistemic Akrasia. *Noûs* 48 (4):718-744.

— Icard, T. Holliday W. Indicative Conditionals and Dynamics Epistemic Logic. (2017). *Proceedings of the Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK XVI).*

— Ichikawa, Jonathan Jenkins and Matthias Steup, (2018). The Analysis of Knowledge. *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.).

— Janssen, Theo M. V. and Thomas Ede Zimmermann.(2021) Montague Semantics. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.)

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

— Joyce, James (2003). Bayes' Theorem. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition).

— Kelly, Kevin (1996). *The Logic of Reliable Inquiry.* Oxford University Press USA.

— Kim, Jaegwon, (1988). What is Naturalized Epistemology? in James E. Tomberlin (ed.), *Philosophical Perspectives,* 2, Atascadero, CA: Ridgeview Publishing Co., pp. 381–406.

— Kornblith, Hillary. (2002). *Knowledge and Its Place in Nature.* New York: Oxford University Press.

— Kremer, Philip and Mints, Grigori. (2005). Dynamic topological logic. *Annals of Pure and Applied Logic* 131 133–158.

— Kripke, Saul (1980). Naming and Necessity. Harvard University Press.

— Lenzen, Wolfgang (1978). Recent work in epistemic logic. Acta Philosophica Fennica 30:1-219.

— Littlejohn, Clayton (2013). No Evidence is False. *Acta Analytica* 28 (2):145-159.

— Lorini, E., and Castelfranchi, C. (2007). The cognitive structure of surprise: looking for basic principles. *Topoi*, 26, 133-149.

— Lyons, Jack, (2016) Epistemological Problems of Perception, *The Stanford Encyclopedia of Philosophy*

— McDowell, John, (1982). Criteria, Defeasibility and Knowledge. *Proceedings of the British Academy.* 68: 455–79.

— McDowell, John. (1994). *Mind and World.* Cambridge, MA: Harvard University Press.

— McKinsey, J.C.C., Tarski, A. (1944). The algebra of topology. Annals of Mathematics (2), 45, 141–191.

— Miller, Brian T. (2016). How to Be a Bayesian Dogmatist. *Australasian Journal of Philosophy* 94 (4):766-780.

— de Montaigne, Michel. Les Essais, (2007) J. Balsamo, C. Magnien-Simonin M. Magnien (eds.) (with “Notes de lecture” and “Sentences peintes” edited by Alain Legros), Paris, “Pléiade”, Gallimard.

— Moretti, Luca and Piazza, Tommaso. (2018). Transmission of Justification and Warrant, *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition).

— Moss, Sarah (2015). Time-Slice Epistemology and Action Under Indeterminacy. In Tamar Szabó Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology.* Oxford University Press. pp. 172–94.

— Mott, Peter (1998). Margins for error and the sorites paradox. *Philosophical Quarterly* 48 (193):494-504.

— Nagel, Jennifer (2013). Knowledge as a Mental State. Oxford Studies in Epistemology 4:275-310.

— Nelson, Michael. Propositional Attitude Reports. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition).

— Neta, Ram and Rohrbaugh, Guy (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly* 85 (4):396–406.

— Okasha, S. (2013). On a flawed argument against the KK principle. *Analysis* 73 (1):80- 86.

— Pritchard, Duncan. (2012) *Epistemological Disjunctivism.* Oxford: Oxford University Press.

— Pryor, James (2000). The skeptic and the dogmatist. *Noûs* 34 (4):517–549.

— Pryor, James (2004). What's wrong with Moore's argument? *Philosophical Issues.* 14 (1):349–378.

— Pryor, James (2013). Problems for Credulism. In Chris Tucker (ed.), *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism.*

— Putnam, Hilary. (1975). The Meaning of Meaning. In *Mind, Language and Reality; Philosophical Papers Volume 2.* Cambridge: Cambridge University Press, 215-271.

— Quine, W.V.O. Epistemology Naturalized. (1969). in *Ontological Relativity and Other Essays*, New York: Columbia University Press.

— Rizzieri, A. (2011). Evidence does not equal knowledge. *Philosophical Studies*, 153, 235–242.

— Rosati, Connie S., Moral Motivation *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.)

— Rott, H. (2004) Stability, Strength and Sensitivity: Converting Belief into Knowledge. *Erkenntnis.* 61, 469–493.

— Roush, Sherrilyn (2017) Epistemic Self-Doubt . *Stanford Encyclopedia of Philosophy.*

— Salow, Bernhard (2018). The Externalist's Guide to Fishing for Compliments. *Mind.* 127 (507):691-728.

— Salow, Bernhard & Ahmed, Arif (2017). Don't Look Now. *British Journal for the Phi -losophy of Science.*

— Sharon, Assaf and Spectre, Levi (2008). Mr. Magoo's mistake. *Philosophical Studies* 139 (2):289-306.

— Sietsma, F. and J. van Eijck. (2012). Action emulation between canonical models, in *Proceedings of the 10th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 10)*, Sevilla, Spain.

— Siegel, S., and Silins, N. (2015). The Epistemology of Perception. In (Ed.), *The Oxford Hand -book of Philosophy of Perception.* : Oxford University Press.

— Stalnaker, R. (1991). The Problem of Logical Omniscience, I. *Synthese*, 89(3), 425-440.

— Stalnaker, R. (2006). On Logics of Knowledge and Belief. *Philosophical Studies.* 128(1)

— Stalnaker, Robert (2015). Luminosity and the KK thesis. in Goldberg, Sanford C. (ed.) *Externalism, Self-Knowledge, and Skepticism: New Essays.* Cambridge University Press.

— Titelbaum, M.G. (2010). Tell me you love me: bootstrapping, externalism, and no-lose epistemology. *Philosophical Studies.* 149, 119–134.

— Troquard, Nicolas and Balbiani, Philippe (2019). Propositional Dynamic Logic. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.)

— Vogel, Jonathan (2000). Reliabilism Leveled. *The Journal of Philosophy.* 97(1): 602–623.

— Vogel, Jonathan (2008). Epistemic Bootstrapping. *The Journal of Philosophy.* 105 (9):518-539.

— Wang, Yanjing (2018). Beyond Knowing That: A New Generation of Epistemic Logics. In Hans van Ditmarsch & Gabriel Sandu (eds.), *Jaakko Hintikka on Knowledge and Game Theoretical Semantics*. Springer. pp. 499-533.

— Wang, Yanjing & Cao, Qinxiang (2013). On axiomatizations of public announcement logic. *Synthese* 190 (S1).

— Weatherson, Brian (2007). The Bayesian and the Dogmatist. *Proceedings of the Aristotelian Society* 107:169-185.

— Weisberg, Jonathan (2010). Bootstrapping in General. *Philosophy and Phenomenological Research.* 81 (3):525-548.

— Weisberg, Jonathan (2012). The Bootstrapping Problem. *Philosophy Compass* 7 (9):597-610

— Weisberg, Jonathan (2015). You've Come a Long Way, Bayesians. *Journal of Philosophical Logic.* 44 (6):817-834.

— Weisberg, Jonathan. (2021) Formal Epistemology. *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.)

— White, Roger (2006). Problems for Dogmatism. *Philosophical Studies* 131 (3):525-557.

— Williamson, Timothy (2000). *Knowledge and its Limits.* Oxford University Press.

— Williamson, Timothy (2008). Why epistemology cannot be operationalized. In Quentin Smith (ed.),*Epistemology: New Essays*. Oxford University Press.

— Williamson, Timothy (2013). Gettier cases in epistemic logic. *Inquiry*, 56, 1–14.

— Williamson, Timothy (2014). Very improbable knowing, *Erkenntnis.* 79, 971–999.

— Williamson, Timothy. (forthcoming) Justification, excuses and sceptical scenarios. Dorsch, F. and Dutant, J.,editors, *The New Evil Demon: New Essays on Knowledge, Rationality and Justification* Oxford: Oxford University Press.

— Williamson, Timothy (forthcoming). Evidence of Evidence in Epistemic Logic. In in Mattias Skipper Rasmussen and Asbjørn Steglich-Petersen (eds.), *Higher-Order Evi -dence: New Essays*, Oxford: Oxford University Press.

— Wittgenstein, Ludwig. (1969). *On Certainty*, Denis Paul and G. E. M. Anscombe (trans.), Oxford: Basil Blackwell.

— Wright, Crispin. (2002). Anti-sceptics Simple and Subtle: G.E. Moore and John McDowell. *Philosophy and Phenomenological Research*, 65: 330–48.

— Wright, Crispin. (2004) Warrant for Nothing (and Foundations for Free)? *Aristotelian Society Supplementary Volume*, 78: 167–212.

— Yap, Audrey, Hoshi, Tomohiro. (2009) Dynamic Epistemic Logic and Branching Temporal Structures. *Synthese* 169 (2009): 259-281.