HOMEOSTATIC EPISTEMOLOGY: RELIABILITY, COHERENCE AND

COORDINATION IN A BAYESIAN VIRTUE EPISTEMOLOGY

By

SUSANNAH KATE DEVITT

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in philosophy

written under the direction of

Ernest Sosa

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

OCTOBER 2013

HOMEOSTATIC EPISTEMOLOGY: RELIABILITY, COHERENCE

AND COORDINATION IN A BAYESIAN VIRTUE EPISTEMOLOGY

By SUSANNAH KATE DEVITT

Dissertation Director: Ernest Sosa

How do agents with limited cognitive capacities flourish in informationally impoverished or unexpected circumstances? Aristotle argued that human flourishing emerged from knowing about the world and our place within it. If he is right, then the virtuous processes that produce knowledge, best explain flourishing. Influenced by Aristotle, virtue epistemology defends an analysis of knowledge where beliefs are evaluated for their truth and the intellectual virtue or competences relied on in their creation. However, human flourishing may emerge from how degrees of ignorance are managed in an uncertain world. Perhaps decision-making in the shadow of knowledge best explains human wellbeing—a Bayesian approach? In this dissertation I argue that a hybrid of virtue and Bayesian epistemologies explains human flourishing—what I term homeostatic epistemology.

Homeostatic epistemology supposes that an agent has a *rational credence p* when *p* is the product of reliable processes aligned with the norms of probability

theory; whereas an agent *knows* that p when a rational credence p is the product of reliable processes such that: 1) p meets some relevant threshold for belief (such that the agent acts as though p were true and indeed p is true), 2) p coheres with a satisficing set of relevant beliefs and, 3) the relevant set of beliefs is coordinated appropriately to meet the integrated aims of the agent.

Homeostatic epistemology recognizes that justificatory relationships between beliefs are constantly changing to combat uncertainties and to take advantage of predictable circumstances. Contrary to holism, justification is built up and broken down across limited sets like the anabolic and catabolic processes that maintain homeostasis in the cells, organs and systems of the body. It is the coordination of choristic sets of reliably produced beliefs that create the greatest flourishing given the limitations inherent in the situated agent.

ACKNOWLEDGEMENTS

DEDICATION

For Morgan

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1   INTRODUCTION

To count as a belief a state has to be part of a system of states in which processes of integration and updating function to keep the subject's mental contents in epistemic equilibrium to some degree or other (Weiskopf, 2008).

1.1   INTRODUCTION

Living creatures exist along a continuum from bare survival to optimized success. Humans survive, satisfice, flourish or optimize depending on their actions at each decision-point. To flourish is to function near optimally, rather than just 'getting by', but optimal functioning depends on useful and reliable information, and, ideally, knowledge.[1] Aristotle was the first western philosopher to explicitly connect human flourishing and knowledge in the *Nicomachean Ethics*:

> Will not the knowledge of [our desires], then, have a great influence on life? Shall we not, like archers who have a mark to aim at, be more likely to hit upon what is right? If so, we must try, in outline at least, to determine what it is, and of which of the sciences or capacities it is the object. (Book.1, §2).[2]

Aristotle's point is that the quest for knowledge is intertwined with the quest to satisfy our desires to achieve *eudaimonia*.[3] This dissertation is a study of the competencies that produce knowledge in the service of human flourishing—an examination of *virtue epistemology*. If Aristotle is right, then it is the virtuous processes that produce knowledge that best explain flourishing. However, it is not clear that Aristotle is right. Human flourishing may emerge from how degrees of ignorance are managed in an uncertain world. Perhaps decision-making in the

---

[1] While knowledge might include abstract ideas about wellbeing, such as the importance of agency, caring for others, mindfulness, eating a balanced diet, daily exercise, wealth (up to a point), routine, self-discipline, and so forth, it mostly refers to simple facts, such as what others are feeling, what the weather is going to be like, and where one might have left their keys. Humans are particularly interested in answers that are tailored to their wants, and in needs that are relevant to their circumstances and objectives. Having attained this knowledge, they then want to put it into practice.
[2] http://ebooks.adelaide.edu.au/a/aristotle/nicomachean/book1.html
[3] *Eudaimonia* is typically translated as "happiness" or "human flourishing" or "wellbeing," and it is an explicitly *rational* satisfaction with life (Hursthouse, 2013).

shadow of knowledge best explains human wellbeing—a Bayesian approach? In this dissertation I argue that both virtue and Bayesian approaches explain human flourishing. Epistemic success is achieved by the deployment of competent processes that produce knowledge and rational credences. What emerges from a hybrid Bayesian Virtue epistemology is a new approach to knowledge and justification I term, *homeostatic epistemology*.

A paradox of knowledge is that humans *think* that they know a great deal—in part, because their lives are filled with success—yet we have ample evidence that we know much less than we think (Kahneman, 2011). A recent headline, "British public wrong about nearly everything" (Page, 2013), illustrates this point particularly well. A survey by the Royal Statistical Society (RSS) and King's College, London (conducted in July 2013) showed that British public opinion is deeply confused about a number of issues, including teen pregnancy rates, benefit fraud, and levels of foreign aid and crime. For example, most people think the crime rate is stable or increasing, when in fact it has been declining for decades. Similarly, the public think that that £24 of every £100 of benefits is fraudulently claimed, even though official estimates suggest that only 70 pence in every £100 is distributed falsely. As the executive director of the RSS concludes, "our data poses real challenges for policymakers. How can you develop good policy when public perceptions can be so out of kilter with the evidence?" (Shah, 2013).

Public outcry against the RSS survey results shows the tremendous scope for confusion and consternation when attempting to interpret statistics and uncover what is *really going on*. Some members of the public (Page, 2013)

criticized the statistical methods, such as the data or sample size (e.g., one respondent claimed that the public somehow knew a fundamental truth about benefits fraud that the establishment could not access), while a commentator with a statistical background said that the results were skewed because the research included pensions—a category with a zero fraud rate. Some attacked the conclusions, arguing that generalized statistics are irrelevant to the 'lived empirical experiences' of each person. In the midst of the furore, a casual observer might reflect on this question: the British public probably knows *something*, but *what* precisely do they know, and, more pointedly, *when* does their ignorance matter?[4] To phrase this slightly differently: What can be known and is knowledge *critical* to flourishing?

So, what can be known? Socrates considered the barriers to knowledge to be so great that he doubted that any *certain* knowledge was achievable (*The Apology* 20e-23c). There are too many reasons why we might be wrong, he argued, for us to be confident that we are right, beyond all doubt, about anything. If Socrates is right, then human flourishing must be dependent on how degrees of ignorance are managed in an uncertain world—the view espoused by *Bayesian epistemology*. Bayesian epistemology and decision-theory avoid the problem of knowledge by studying rational degrees of belief, given limited evidence and cognition. Bayesians point out that, given the boundless ignorance in our day-to-day lives,

---

[4] Having knowledge is important to flourishing, but is all knowledge equally important? For example, if the British people are ignorant on national issues such as immigration, welfare, foreign aid, and so on, then false beliefs drive their vote. This means that policies commensurate with these false beliefs will be pursued by politicians seeking office, regardless of objective benefits to the public. On the other hand, the British people may know how to improve their benefit payments through loopholes and manipulation of the system, which vastly improves their weekly income and helps them and their families live better lives. Clearly, when it comes to flourishing, some knowledge seems less important than other knowledge. Philosophers (Pritchard & Turri, 2012) have argued that accumulating true beliefs independently of their prudential value is of little value. I explore the value of knowledge in *Chapter 3*.

certainty cannot be a prerequisite for action or success. Still, does this mean that the knowledge baby should be thrown out with the exacting bath water? Weighing up Aristotelian and Socratian approaches, the central question of this dissertation is: How do cognitively limited agents flourish with limited capacities in informationally impoverished circumstances? Homeostatic epistemology is the thesis that agents flourish by aligning their confidence with available evidence and relevant circumstances, and that a hybrid of virtue and Bayesian epistemologies (BE) justifies this belief alignment.

Both Bayesian and virtue epistemologies contend that competent processes require *reliability* (consistent true belief formation) and *cohesiveness* (beliefs mutually support one another). The problem with reliability is that, in isolation, it struggles to explain reflective knowledge. BVE offers an iterative reliability to justify reflective knowledge. However, BVE cannot resolve a central problem of cohesiveness, namely, that increasing the number and cohesion of beliefs may paradoxically make them more likely to be false.

I argue that this paradox arises because coherence is typically evaluated holistically (across all beliefs), rather than within limited sets or partitions. I argue that limited coherence resolves traditional concerns and raises the need for another kind of justification, *coordination* (integration across sets of beliefs). Homeostatic justification is a reliabilist account of foundational, cohering and coordinated beliefs.

In this chapter, I lay the groundwork for the chapters to come. In §1.2, I introduce the biological concept of homeostasis and apply it to epistemic

processes and outcomes. I also lay down the central commitments of homeostatic epistemology. In §1.3, I argue that naturalized epistemology qua cooperative naturalism is the most promising framework with which to approach the study of knowledge and decision theory. Cooperative naturalism is the thesis that philosophical findings can shed light on empirical science and that, conversely, empirical science can influence disputes in traditional philosophy. In §1.4, I introduce reliabilism, which is promising because knowledge can be objectively evaluated through empirical research on cognition and performance success in the world. In §1.5, I define reliabilist justification grounded in empirical research on memory systems. I go on to articulate two prominent and contemporary variants of reliabilism: virtue epistemology and Bayesian epistemology. In §1.6, I discuss virtue epistemology, which defines knowledge as reliable beliefs brought about by the internally driven competencies of the agent. In §1.7, I explain how Bayesian epistemology circumvents the problem of knowledge by focusing on credences constrained by probabilistic laws, and that a hybrid of virtue and Bayesian epistemologies is the foundation for answering how both knowledge and justified beliefs (which fall short of knowledge) are valuable to human flourishing. In §1.8, I conclude this chapter by highlighting some of the problems associated with a hybrid account, outlining how these problems will be addressed in subsequent chapters, and justifying the emergence of homeostatic epistemology.

1.2   Homeostatic epistemology

> The integrating power of the nervous system has in fact in the higher animal, more than in the lower, constructed from a mere collection of organs and segments a functional unity, an individual of more perfected solidarity. (Sherrington, 1906, p. 353)

Homeostasis refers to self-regulating physiological processes that maintain the stability of the body's internal systems, including metabolism, blood pressure, and body temperature.[5] Homeostasis is a coordinated response to ensure internal stability (Denton, McKinley, & Wessinger, 1996), and it has two key features: 1) an optimum state for an organism (e.g., a 37 degree core internal temperature), and 2) processes that enable the organism to strive for and maintain this optimal point (e.g., sweating, shivering). The optimum state can vary depending on context. For example, seasonal differences alter the optimum temperature. Homeostatic processes are both reactive and predictive (Moore-Ede, 1986; Mrosovsky, 1990; Wingfield & Ramenofsky, 1999). For example, rats build a nest in response to sudden cooling in their environment (Richter, 1943), but they also predict colder temperatures with circadian timing systems (Gallistel, 1990) and secrete insulin when they perceive food sources. Homeostatic processes are also not confined to local physiological function, but instead are orchestrated by the central nervous system and/or placenta. For example, the hormone CRH[6]

---

[5] Since the late 1980s, another term—'allostasis'—has entered the lexicon regarding body regulation. Allostasis refers to the body's capacity to anticipate needs, while homeostasis is defined as being purely reactionary. Processes described by advocates of allostasis include lactation and migration, both of which involve a complex array of physiological changes (including gene expressions) to ensure wellbeing (McEwen & Wingfield, 2010). Allostatic processes have shifting equilibrium points (rather than one ideal) and they differ between individuals, as opposed to being bound by species-specific parameters. Critics of the term argue that it mischaracterises homeostasis and offers no explanatory benefits (Day, 2005; Romero, Dickens, & Cyr, 2009). Cannon's original concept of homeostasis allowed for both anticipatory functions and wide-ranging parameters on many physiological systems, and for this reason I use the term homeostasis to refer to both allostatic and homeostatic processes. The differences between the two terms are not pertinent to this thesis, and a broader use of 'homeostasis' is consistent with the literature (Nelson & Drazen, 2000). For further discussion on this topic, see (Ganzel, Morris, & Wethington, 2010; LeDoux, 2012; McEwen & Wingfield, 2010; Romero, et al., 2009; Schulkin, 2003).
[6] Corticotropin-releasing hormone (CRH) is a peptide hormone and neurotransmitter that manages the internal response to stress including the release of adrenocorticotropic hormone (ACTH) and cortisol. The release of ACTH and cortisol set off

coordinates a range of processes that dictate normal or preterm birth (in the case of severe maternal stress) (Schulkin, 2003).

The objective of homeostasis is to maintain equilibrium against changing internal and external conditions. Consider an underdressed shift-worker stepping outside at night to watch snow tumble silently onto an empty street. Temperature receptors on her skin transmit information to the hypothalamus, alerting her to the sudden change in temperature. The hypothalamus then directs effectors (sweat glands and muscles) to increase heat within the body to ensure a core temperature of 37 degrees Celsius. If the worker is unable to internally regulate her temperature or alter her environment (e.g., by seeking shelter or a jacket because the security door shuts and she has left her keys inside), the body goes into survival mode, withdrawing heat from the extremities (vasoconstriction) in order to keep core organs functioning. The longer the body remains in sub-optimal conditions, the more severe the deficits to functioning, including confusion, memory loss, exhaustion, difficulty moving, severe shivering, slurring of speech, and eventually cardiac arrest. However, when the worker manages to slide open a window and return once more to her warm office, the body quickly senses the shift and begins returning the blood supply to normal operations, and normal cognitive operations resume. A creature *flourishes* when physiological processes are in equilibrium, but they can *survive* within a broad range of conditions.

Drawing from homeostatic processes, homeostatic epistemology refers to the processes that maintain rational beliefs across changing internal and external

---

many systems including immune responses, the regulation of inflammation, carbohydrate metabolism, protein catabolism, blood electrolyte levels, and behaviour (Rogers, 2012).

information environments. For example, take a tourist observing a hand gesture upon arriving in a foreign country. This gesture connotes a friendly welcome at home, yet it is accompanied by oddly distracted body language in this instance. What should the tourist believe? The tourist must evaluate his belief about the gesture alongside other beliefs about the culture, circumstances, and the individual he is observing. Responding too warmly might provoke suspicion in a customs officer or exhausting provocation from a manipulative trader. Withholding emotion, however, might result in a missed opportunity for an exquisite local meal or a lost opportunity to help someone in genuine distress. We flourish when we pick the right degree of belief, given our information and our goals, and when we act in correspondence with that belief.

Historical precedent for homeostatic epistemology can be found in the work of Charles Sherrington, who described the brain as "an enchanted loom where millions of flashing shuttles weave" (Levine, 2007; Sherrington, 1906). Born out of the surge of knowledge in metabolic neurochemistry at the turn of the 20th century, the 'enchanted loom' metaphor describes the interlacing of electrical and chemical neural components to produce mental activity. Sherrington supposed that the mind consists of many systems coordinating a single tapestry, rather than a central or single soul directing the weaving process. His use of the word 'shuttle'— an object that moves backwards and forwards across the threads—may be an attempt to convey the importance of oscillating anabolic and catabolic processes to cell life and, by extension, the influence of balancing forces in the

mind more generally. Nerve cells, for example, were intriguingly both independent and interactive through the action of hormones.[7]

Sherrington had two key ideas that are pertinent to homeostatic epistemology: 1) the compatibility of independent and interacting processes, and 2) anabolic and catabolic processes at multiple levels of biological function. Translated into epistemic notions these ideas become: 1) sets of beliefs are justified by lower level processes and goals *and* by the coordination of processes to achieve higher level goals; and 2) the justification of a belief varies by its relationship with other beliefs. Greater coherence is built up (anabolic) by the number of beliefs and their logical relationships. However, justification reaches an equilibrium point to achieve an epistemic goal, a point of maximal fittedness between beliefs after which time more beliefs do not increase justification. At times justification increases by breaking down (catabolic) large sets of beliefs into smaller partitions relevant to more specific goals.

Agents flourish when they *strive* for goals that are attuned to ecological context, and when they *adapt* their thoughts, behavior, and goals in challenging contexts. For example, a paleontologist in the field succeeds not only when she relies on her knowledge of paintbrushes, dental picks, and plaster to remove fossils, but also when she adapts techniques in challenging circumstances (for example, using toilet paper to cast a crocodile skull when atypical flooding inundates a fossil site, rendering plaster as useless as mud). Similar fossil rescues

---

[7] The effect of glands such as the adrenal and pituitary on the nervous system reduced the brain's pre-eminence as the controlling organ of the body's state. Sets of processes worked in functional solidary to maintain the wellbeing of the body. Though inspired by Sherrington, I do not wish to engage in a philosophical discussion about whether the mind has a central processor (or soul) or whether the mind consists of encapsulated modules (Fodor, 1983).

by untrained enthusiasts in either normal or aberrant circumstances could result in irreparable damage to the fossil record. Agents flourish when they detect the limits of their own capacities to achieve appropriate aims, and knowledge occurs when circumstances and capacities align to produce optimal outcomes for the agent.[8]

The Socratic standard for certain knowledge is too austere to make sense of success. Instead, knowledge ought to be contextualized in terms of the agent's goals and conditions. Additionally, and contrary to traditional epistemic claims, human flourishing does not always require knowledge, and beliefs forged through processes that are 'good enough' can yield satisfactory outcomes[9]. Take losing one's keys as an example: When we realize that our keys are not in our bag or in the little wicker basket on the bookshelf, we do not have to *know* where they are to start being productive in our search for them. A set of hunches are good enough most of the time—perhaps the keys are still in the front door? Could I have left them on the kitchen bench when I brought home the shopping? Could they be in the jacket I wore recently? Even when we are in a hurry, our rational credences in probable hypotheses help narrow the scope of where our keys may be. These rational credences are justified with Bayesian apparatus.

---

[8] How is knowledge obtained? How does a palaeontologist gain knowledge and capacities? Traditional epistemologies differentiate between innate, rational knowledge and empirically grounded knowledge that one has learned. Plato thought that humans were innately endowed with abstract knowledge, such as arithmetic axioms or geometric forms (e.g., the concept of a triangle). Descartes argued that innate intellectual powers, rather than the senses, could discern the truth. In this dissertation, I follow in the Western analytic empiricist tradition of Aristotle, Locke, Hume, and Berkeley to study learned knowledge. Humans learn through experiences in the world. Higher level knowledge must somehow be scaffolded from perceptual, memorial, and testimonial sources. Knowledge arises from the operation of contextually appropriate, reliable processes, which produce specific cognitive outcomes. Key mechanisms for belief revision include causal Bayesian processes, which are the subject of *Chapter 2*.

[9] Of course, it that it might be that *overall* human flourishing does always require quite a lot of knowledge, even though there are items that aid human flourishing, such as finding your keys that do not require knowledge (E. Sosa, personal communication, August 5, 2013)

I address a number of arguments that can be made against Bayesianism: 1) that Bayesian models can be made to fit almost any data, 2) that Bayesian psychology is incomplete, and 3) that Bayesian norms are impossibly high. In §1.7, I argue that Bayesian norms are the only guide we have to rational decision-making, but reliable human processes are shaped by Bayesian norms, as well as other factors such as causal reasoning. Bayesian processes require reliability (consistent true belief formation) and cohesiveness (where beliefs mutually support one another).

Reliabilism best accounts for low-level, mechanistic, cognitive processes, such as the habituation and association characteristic of animal knowledge, but how can low-level processes that produce animal knowledge explain the production of high-level reflective knowledge? In *Chapter 2*, I argue that iterative and hierarchical mechanisms (e.g., causal Bayes nets) can account for reliably produced, high-level reflective knowledge. However, Bayesian processes are underdefined, and an additional measure, such as coherence, is needed to bolster Bayesian reliabilism.

Cohesiveness between beliefs involves logical consistency, which ought to help Bayesians objectively assign prior probabilities to hypotheses. The problem with cohesiveness, however, is that adding more logically consistent beliefs to a set of beliefs paradoxically makes the set more likely to be false—a threat to epistemic holism. In *Chapter 3*, I respond to this with the thesis that coherence is valuable between *limited* sets of beliefs, which I term *chorism*.[10]

---

[10] chorism from the greek chórisma(χώρισμα) meaning 'partition'

Chorism refers to the view that epistemic justification is not about a single justified true belief (*atomism*), nor an integrated holistic web of beliefs (*holism*), but instead is about the value of a subset of mutually supporting beliefs that are relevant to an agent's success. It is the coordination of *choristic* sets of cohesive beliefs that create the greatest flourishing, given the limitations inherent in the situated agent. Choristic sets of belief explains instances of knowledge where the agent may hold inconsistent beliefs and also explains how competently produced beliefs that fall short of knowledge are, nonetheless, vital components of a thriving agent responding to a complex and intractable world.

Homeostatic epistemology argues that instead of a holistic web of belief, the mind is *choristic*, consisting of coordinated partitions of mutually supporting credences and beliefs. Organisms achieve knowledge through the combined action of integrative and disintegrative cognitive processes, mirroring the anabolic and catabolic forces in the body. Epistemic homeostasis occurs when epistemic objects are coordinated to produce the greatest flourishing. Reflective knowledge represents the highest form of flourishing *eudaimonia*.

1.3   Naturalized epistemology

> As to this last, Natural Science in its progress can provide a frame of reference for Natural Theology. . . . Natural Science is a branch of knowledge by general consent not primarily based on the a priori. It derives essentially from details. It amasses them and lives on and by them. Its generalizations so built up are even when arrived at constantly being controlled by fresh details. In that way its generalizations do from time to time suffer change. Natural Theology is interested in this as a background and context for its own text. (Sherrington, 1941)

Charles Sherrington argued that because mind and body emerged from the same physical substrate, insights formed by the mind are constructed and directed

by the same processes that affect the external world. As a result, natural science is a constant process of renegotiation between scientists and philosophers, and this dissertation approaches epistemological questions using both philosophical considerations and empirical results. I treat epistemology as a naturalized project (Devitt, 1997 §5.8; Kornblith, 1985; Quine, 1969) because our best explanations for how we know are informed by methods and findings from successful scientific practice.[11]

Naturalized epistemology seeks empirically grounded explanations for epistemic concepts such as justification, evidence, and normative practice. Naturalized epistemologists use initial intuitions to contemplate epistemic issues, but they do not take them for granted.[12] Thus, an inferential principle or process is validated based on how it performs in the world. Basic truths become known via perceptual or sensory processes; the concept 'perpendicularity,' for example, is established by postnatal exposure to both vertical and horizontal properties (Blakemore & Cooper, 1970). Investigating our cognitive architecture and environments will shed light on how experience creates abstract ideas, and on how valuable these ideas are in terms of our success.

Naturalized epistemology is unaffected by the diversity of intuitions regarding epistemic concepts because they do not take intuitions as decisive

---

[11] I am also a realist about truth embracing a correspondence theory (Devitt, 1997 §6.6; Putnam, 1978, pp. 99-103). The correspondence theory argues that the best explanation for an individual's capacity to meet their needs and desires is that there is a faithful correspondence between what is believed and what actually is. That is, "an organism's belief that is true is conducive to the fulfilment of its desires, and one that is false is conducive to their frustration" (Devitt, 1997, p. 99). True beliefs lead to the fulfilment of anticipated experiences.

[12] Bealer (1987, 1993, 1998) argues that naturalized epistemology cannot avoid the a priori because it depends on inferential principles that are derived through rational intuition. Kornblith (2002) responds that internally generated principles are justified by naturalists due to their reliability not because they arose via the intellect. I assume Kornblith's view is the right approach.

evidence for the sufficiency of epistemic theories[13]. The object of epistemology is not the study of epistemic concepts, but rather the success of thinking creatures in the world.

One way to conceive the relationship between traditional and naturalistic epistemological projects is to assert a supervenience relationship (Kim, 1988). That is, we can suppose that epistemic properties supervene on non-epistemic, naturalistic conditions. If it turns out that humans abide by Bayesian hypothesis testing in a range of conditions, then the normative rules established in probability theory become normative guidelines and supervene on the agent in situ. Similarly, if it turns out that humans behave differently depending on how they assent to propositions (i.e. whether they 'suspect', 'believe' or 'know' that p), then epistemic principles regarding degrees of belief and knowledge will supervene on their appropriate or ideal use. Bayesian principles formalize justification in scenarios in which people assent to propositions that they "know well" or they "really know" rather than just "knowing"; or that they "are pretty sure about," "don't know" or "doubt completely." Bayesian principles also make sense of why testimony expressed with degrees of confidence offers evidential support.

Supposing that the descriptive naturalistic project (e.g. modeling, simulation, experimentation, cross-cultural analysis) is helpful in specifying how we come to the views we have, it may be unclear how epistemic concepts,

---

[13] Traditional epistemic methods (both naturalist and rationalist) have been empirically challenged with philosophical experiments (Knobe & Nichols, 2008). Methods under scrutiny include: linguistic or conceptual analysis, intuitive strength, clarity, and so forth. For example, experimental philosophers have found intuitions of knowledge on Gettier cases vary depending on cultural, educational and linguistic background. (Weinberg, Nichols, & Stich, 2008). Still, one does not need to experiment to find conflicting intuitions. Highly trained western, analytic philosophers use these methods and disagree on the answers (see philosophical disagreements regarding Gettier cases such as, Hetherington, 1999; Klein, 1976; Turri, 2012)—suggesting that the methods themselves do not produce objectively likely results. If some of these varied responses are objectively correct, and some are objectively incorrect, then it is unclear what sort of defensible theory of traditional epistemic methods could explain such divergence (Weinberg, et al., 2008).

particularly normative epistemic concepts about justification or knowledge remain relevant to epistemic inquiry. What do the reflective intuitions of the 'armchair epistemologist' inform? Cooperative naturalism attempts to save the armchair.

Cooperative naturalism (Feldman, 2012) supposes that problems and solutions identified in traditional epistemology ought to inform and shape empirical and mathematical work in cognitive science and, conversely that epistemologists ought to consider empirical experiments and rigorous models by cognitive scientists (psychologists, anthropologists, neuroscientists, and so on) rather than depend exclusively on their naïve conceptions of human functioning. This is not a new view. In the science-to-philosophy direction, Ewing (1934) argued that "it would… fall to the studies of cognition to specify just which processes reliably engender true belief and how, thus giving substance and strength to our overall theory of knowledge" (p.64). In the philosophy-to-science direction, cognitive scientists use philosophical arguments about cognitive architecture (Fodor & Pylyshyn, 1988) to evaluate empirical models (Bowers & Davis, 2012). [14] I take on the cooperative naturalist position. Traditional philosophical methods are particularly helpful for the normative project, but they can perform clarificatory work for the descriptive project too.

1.4 RELIABILISM

Skeptical arguments show that there are no necessary deductive or inductive relationships between our beliefs and their evidential grounds or even their probability (Greco, 2000). In order to avoid skepticism, a different view of what

---

[14] Other non-empirical philosophies influence cognitive science including research into pure mathematics, modeling and logic.

constitutes good evidence must be found. Good evidence for a positive epistemology might be the reliable connection between what we believe about the world, and the way the world behaves (which is consistent with these beliefs), such that "the grounds for our beliefs are reliable indications of their truth" (Greco, 2000, p. 4). Reliabilism supposes that a subject knows a proposition *p* when a) *p* is true, b) the subject believes *p* and c) the belief that *p* is the result of a reliable process. A key benefit of reliabilism is that beliefs formed reliably have epistemic value, regardless of whether an agent can justify or infer reasons for their reliability. Reliable beliefs, like the readings from a thermometer or thermostat, are externally verifiable.

Cognitive agents are more complex than thermometers, however. Agents have higher-order reliability based on the reliability of sub-systems. In complex agents, the *degree* of reliability of the process gives the *degree* to which a belief generated by it is justified (Sosa, 1993). The most discussed variant of reliabilism is process (or paradigm) reliabilism:

> Process reliabilism: *"S's belief in p is justified iff it is caused (or causally sustained) by a reliable cognitive process, or a history of reliable processes"* (Goldman, 1994).

Take the capacity to recognize faces as an example. There are many processes involved with face recognition, ranging from low-level perceptual processes regarding the contours, textures and features of the physical face; to memory processes that trigger a sense of familiarity; and high-level recognition processes that identify the face as being that of a particular person. Individuals vary in how reliably they recognize faces based on how reliably each of these processes operate. When these processes are disjointed, people can fail to

recognize faces they know (prosopagnosia) or they may recognize faces, but nonetheless believe that person to be an imposter (Capgras delusion). Prosopagnosics show no explicit face recognition, although implicit tests—such as skin conductance response—may reveal that they do retain recognition at a lower level (Young, 2008). Individuals suffering from Capgras delusion have no difficulty recognizing faces, but they may claim that their spouse or loved ones are imposters because they lack appropriate feelings of familiarity (Gerrans, 1999). The reliability of facial recognition is tied to the reliability of each of the processes underlying it and the degree to which each process functions[15].

A significant issue for reliabilists is a lack of sophistication about how cognitive processes actually operate. For example, Feldman (1985) uses the process type 'inferring' to discuss the implications of categorizing all instances of inference under the same justificatory process, because, as he rightly points out, some inferences are more justified than others. He uses the granularity of inference to argue that reliabilism fails to capture the nuance of ordinary reasoning. However, treating 'inference' as a single process shows naivety given the complexity of cognitive processes.

There are many cognitive processes involved in creating true beliefs. To discuss the success or failure of reliabilism, one must understand the basic principles by which humans come to know the world—i.e. through experiences and the retention of those experiences. The systems responsible for the retention

---

[15] One of the biggest problems for process reliabilism is that a reliable belief may be produced by a multiple processes, each with varying degrees of reliability such that there is not a one-to-one relationship between the reliability of process and the reliability of belief (Conee & Feldman, 1998). This has become known as the Generality Problem. Because some commentators (Bishop, 2010) have argued that the generality problem exists for *all* theories of justification (not just reliabilism), I have not attempted to resolve it here.

of experience are the systems of memory, and they are discussed in the section below.

1.5 SYSTEMS OF MEMORY

> Please assume… that there is in our souls a block of wax, in one case larger, in another smaller, in one case the wax is purer, in another more impure and harder, in some cases softer, and in some of proper quality… Let us, then, say that this is the gift of Memory, the mother of the Muses, and that whenever we wish to remember anything we see or hear or think of in our own minds, we hold this wax under the perceptions and thoughts and imprint them upon it, just as we make impressions from seal rings; and whatever is imprinted we remember and know as long as its image lasts, but whatever is rubbed out or cannot be imprinted we forget and do not know (Plato, 2012)

Memory is a consequence of the fundamental plasticity of the brain (Eichenbaum, 2008). Memory is neither a localized storehouse nor a globally distributed network of information. Instead, memory ought to be considered a fundamental property of brain systems responding to experience. A memory system is any biological process that retains an impression from experience over time. Anything learnt by an organism that affects subsequent thoughts or behavior is a product of a memory system. Memory is both reconstructive and creative, using a range of processes or mechanisms to store, retrieve, calibrate, recognize, construct and respond to previously encountered information. The number and function of memory systems is still only partially understood.

The generally accepted anatomical divisions of memory include: perceptual memory (striatum cerebellum), working memory (hippocampus), procedural memory (brainstem and spinal motor outputs), emotional memory (hypothalamus autonomic and endocrine outputs) and declarative memory (semantic and episodic recollection) interacting with the amygdala and cortical association areas (Eichenbaum, 2008, p. 5). Processes that drive memory include: habituation,

muscle memory, association, recognition, recollection, and reconstruction, as well as processes traditionally linked to the intellect such as hypothetical reasoning and inference (deductive, inductive and abductive) (Tarricone, 2011). Although, memory systems can work independently, they usually work in concert; generating beliefs, feelings and reactions in the agent. The justification of a belief depends on understanding the range of processes involved in its generation. There are two broad categories of memory I focus on in this section—implicit and explicit[16].

1.5.1 IMPLICIT MEMORY

Implicit memory systems are grouped as those that operate beneath the level of conscious awareness or reflection, regardless of their underlying representational structure (Schacter, 1987). These systems are also referred to as 'non-declarative' because their content cannot be declared or verbalized. Implicit memories are learnt without necessarily any planning or intentionally attempting to acquire information or abilities. That is, an organism's learning environment can be used to predict the formation of regularities in behavior without explanatory recourse to mental states. While mental states can reflect upon implicit processes, those mental operations are explicit, not implicit. There are four major categories of implicit memory: habituation, association, skills and implicit knowledge. Of these, this research only focuses on implicit propositional knowledge.

---

[16] There are also numerous minor memory systems defined by the part of the body involved in retention or the body system it services. For example, trans-saccadic memory is the process of retaining information across a saccade—fast, micro-eye movements that build up a 3-D representation of a visual scene (Henderson, 1997). The enteric nervous system shows non-pathological adaptive changes in response to altered digestive activity (Furness, Clerc, & Kunze, 2000).

1.5.1.1 HABITUATION

The most basic sort of implicit memory is habituation or sensitization, where an organism's systems adapt to environmental stimuli but not as a result of motor fatigue or sensory adaptation. For example, plants adjust the angle of their leaves in response to the quality of light, a sea slug retracts its body upon being prodded, or a person becomes oblivious to background noise after a period of time in a noisy location. Habituation is a short-term response to immediate causal influences in an organism's experience constrained by biological parameters. Visual illusions—such as seeing a red afterimage on a white wall after staring at a green square—are the product of the habituation of neurons, which cease firing after prolonged stimulus and then leave a 'shadow' when attention shifts to a neutral background.

1.5.1.2 ASSOCIATION

Association enables an organism to make inductive decisions for the future based on a situation's similarity or contiguity with prior experience. James Mill (1878) gives the following example:

> When, looking at a river, we pronounce its name, we are properly said to exemplify contiguity; the river and the name by frequent association are so united that each recalls the other. But mark the steps of the recall. What is strictly present to our view is the impression made by the river while we gaze on it. It is necessary that this impression should, by virtue of similarity of identity, re-instate the previous impression of the river, to which the previous impression of the name was contiguous (p.121).

When responses to stimuli are retained within an organism, affecting subsequent reactions to similar provocation, an association is said to have formed. Stimuli are associated via temporal proximity (contiguity), syntactic or semantic

similarity and affective resonance, and associations are strengthened via repetition and positive or negative reinforcement (Palmer, 2010).[17] Theories of association in the explanation of memory go back at least as far as Plato. In the Phaedo (*Meno* 82E), Socrates describes recollection as the process of one thing putting you in mind of another[18]. Aristotle similarly argues that recollection involves a succession of associated ideas (DM 451b10-452a7).

Associationist theories of memory were explored and refined by numerous enlightenment philosophers, including John Locke (1689b), David Hume (1739b), and John Stuart Mill (1882). Hume thought that association was the fundamental cognitive act that turned experience into ideas, through which we establish beliefs about abstract notions of identity and causation.

> The nature of experience is this. We remember to have had frequent instances of the existence of one species of objects; and also remember, that the individuals of another species of objects have always attended them, and have existed in a regular order of contiguity and succession with regard to them (Hume, 1739a).

The empiricists considered how various factors influence the strength of associations formed. Three types of influences were postulated: *vivacity* (or distinctness), *duration* (study time) and *frequency* (repetitions) (Price, Laird, & Wright, 1936). Theories were also posited about what would influence the retrieval of associations. It was hypothesized that the degree of resemblance to the stimulus and the recency of the learning event would impact association, as would the absence of interference or being in an altered state of mind[19]. Contiguously

---

[17] James Mill (1878) also included contrast as a law of association, as in habitual pairs of concepts: light and dark, heat and cold, up and down, life and death. He notes that these pairs are always presented because individual concepts are defined relationally contrasted to the other.

[18] Note that the use of 'recollect' does not mean the same in Plato's usage as in modern western considerations. Plato uses the Greek verb *anamimneskesthai*, which more accurately means 'to be reminded' (Sorabji, 2004, p. 35).

[19] Interference refers to the "the coexistence of fewer alternative associates to the cue" (Bower, 2000, p. 4). Temporary diversities of state included intoxication, delirium, depression and so on.

presented words or words with similar meanings (semantic associates) are also associated, meaning that if one is recalled, associated words are also brought to mind (Polyn, Kahana, & Norman, 2009).

Perhaps the most famous experimental example of association is Pavlov's (1927) dog. In this case, the dog learnt the association between the dinner bell and subsequent food, such that he would salivate merely upon hearing the bell. More technically, a conditioned stimulus (e.g., a sound) predicts the future arrival of the unconditioned stimulus (e.g., food), triggering a future-directed response (e.g., salivation). In operant conditioning, an organism's behavior predicts a certain outcome (reward). Unlike habituation, associations are formed, maintained, and altered throughout an organism's life via plastic neural pathways and adaptive biological systems. While associations can be formed in a single instance— contrary to Behaviorist rhetoric (Gallistel, 1990)—competence at various skills increases through repetition.

1.5.1.3 PROCEDURAL KNOWLEDGE

Amongst philosophers, skills or capacities like riding a bicycle, catching a ball, or touch typing are referred to as 'procedural knowledge', 'habit memory' and 'knowledge-how' (Ryle, 1949). Procedural knowledge can also be referred to as embodied memory, such as the muscle memory retained by professional athletes (Sutton, 2006). The use of the word 'knowledge' is misleading, however, because it suggests a degree of representation that may or may not be present depending on the skill in question. Traditionally, philosophers also speak of procedural memory as 'dispositional memory.' Dispositions are a tendency to

behave in a particular way in certain circumstances or an ability that lasts within an individual.

1.5.1.4 REPRESENTATIONAL KNOWLEDGE

The most sophisticated implicit memory is representational. Organisms are capable of creating, retaining, and utilizing vastly complex models of the world to navigate their lives without necessarily having conscious awareness, propositional beliefs or higher order thought. Representational implicit memory includes: cognitive maps (Tolman, 1948); magnetoreception, which allows birds to orient themselves using magnetic fields (Mouritsen & Ritz, 2005); and squirrel caching behavior (Kamil & Gould, 2008).

Some implicit representational memory is propositional, meaning that it can be language-like (Fodor, 1975; Fodor, 2008) or map-like (Blumson, 2012) exhibiting systematicity and productivity. Implicit propositional beliefs are sometimes referred to as 'tacit knowledge' (Fodor, 1968a; Lycan, 1988) or 'unconscious recollection' (Ayer, 1956)[20]. While some implicit beliefs remain hidden from conscious reflection or top-down processing, a subset of implicit beliefs are 'background beliefs' that can be either implicit or explicit depending on attention. For example, a background belief that climate change is real might influence the occurrent belief that a waterfront property is expensive to maintain.

Background beliefs work in the 'background' of an agent's cognitive architecture, but they can nonetheless become the focus of attention or scrutiny.

---

[20] Though there is a good deal of debate whether tacit knowledge is propositional or not (Fantl, 2012), I do not consider the issues here. It is sufficient for my purposes *that* some implicit knowledge is propositional. That is, that some beliefs are not immediately available to consciousness

Occurrent beliefs are beliefs that are being attended to by an agent, but background beliefs can become occurrent beliefs and vice versa, depending on an individual's focus or interest.

1.5.2 EXPLCIT MEMORY

In contrast to implicit memory, explicit memory requires organisms to be capable of second-order thoughts about first-order mental states and it is propositional in a way that implicit memory need not be. Explicit memory is also more flexible than some implicit memory because it is responsive to top-down influence (rather than simply involving bottom-up reactions to perceptual stimuli).

Explicit memories are decoupled from perception, and thus are able to be considered and utilized for problem solving or reconsidering the world. Explicit memory is typically divided into two types: semantic and episodic (Tulving, 1972)—a distinction that is supported by experimental evidence, brain injury impairments and imaging studies.[21] For example, an amnesiac patient might know that the capital of France is Paris, or how to tie his shoelaces, yet may be unable recall any events that lead to this knowledge (Rosenbaum, Murphy, & Rich, 2012). Similarly, a patient might retain intelligence, language and reasoning ability, even with severe anterograde amnesia for information encountered post-injury (Scoville, 1968).

---

[21] The neural correlates of episodic encoding and semantic retrieval "are dissociable but interact in specific brain regions" (Prince, Tsukiura, & Cabeza, 2007).

1.5.2.1 SEMANTIC MEMORY

Semantic memory is the faculty that processes particular experiences into abstracted, more generalized knowledge of the world (e.g., that 5 x 7 = 35) and it is the basis of inferential and analogical reasoning. Semantic memory does not require the agent to mentally time travel to the occasion of learning, but it does mean that they can explicitly use this knowledge. Participants can know details of a prior event without 'putting themselves in the past.'

Semantic memory also does not require conscious deliberation, as even sleep transitions the experience of particulars into generalities, universals and abstractions (Reid & Gaskell, 2012). As Payne et.al. (2009) point out, sleep does not merely consolidate memories, it also transforms and restructures them so that insights, abstracts, inferences and integration can occur (p.333). Empirical findings are contrary to Aristotle (Posterior Analytics) who thought the act of 'demonstration' required conscious efforts.

Philosophers have called semantic memory 'factual-memory'. Ayer (1956) describes it in the excerpt below:

> Suppose that I am set to answer a literary questionnaire, and that I have to rely upon my memory. I shall, perhaps, succeed in remembering that such and such a poem continues in such and such a way, that So-and-so was the author of such and such a book, that a given incident appears in this novel rather than in that. But none of this need involve my having any recollection of a past event. I may recall some of the occasions on which I read, or was told about, the books in question, but equally I may not (136-137).

Ayer considers semantic memory the height of memory function. He supposes that the need to reflect on or entertain memory images reveals that the subject matter is less well known that it optimally would be in the circumstances.

The phenomenal characteristics of memory, such as imagery, come from another explicit memory system—episodic memory. While semantic memory retains facts and generalizes from experience; episodic memory maintains particulars of past experience—including abstract or conceptual facts about the past (e.g. "I've had a car for six months", or "The Iraq war started when my grandmother died") and event specific knowledge (ESK) that includes sights, sounds, smells, feelings and thoughts from the remembered event (Mitchell & Johnson, 2009).

1.5.2.2 EPISODIC MEMORY

Episodic memory draws on many implicit and explicit memory systems. The defining feature of episodic memory is that it can accurately identify a representation as a memory of a particular past experience.[22] As Aristotle says, "whenever someone is actively engaged in remembering, he always says in his soul in this way that he heard, or perceived or thought this before" (449b22-24). Similarly, Locke believed that "…the mind has a power in many cases to revive perceptions that it has once had, with attached to them the additional perception that it has had them before" (1689a). Episodic memory enables us to reflect on the past, untangle the present as well as plan for the future. It is defined by three properties:

1. the retention of information (what, when, where), phenomenology and associated thoughts from an event,

---

[22] Unlike Hume's claim that memories have a particular 'force and liveliness,' episodic memories range from highly vivid, detailed recollections (e.g. flashbulb memory or verbatim trace) to vague feelings of knowing (e.g. remembering the gist of a movie plot or experiencing a sense of familiarity when seeing a peer at a conference) (Brainerd & Reyna, 2002; Jacoby, 1991; Luminet & Curci, 2008; Reyna & Brainerd, 1995)

2. the subsequent re-experience of information, phenomenology and thoughts upon retrieval and

3. the correct identification during retrieval of the re-experienced content as being *of* that event.

The recollection of episodic memories activates brain regions originally involved in the learning episode—such as emotions, mental images, smells or the 'sense of being there'—and the output from these different modalities and processes combine to constitute an episodic memory (Johnson, 2006). Analytic philosophers have long suggested that some of the same parts of the mind activated during perceptual experience are reactivated during episodic memory retrieval.[23] In the 20th century, the cognitive neuroscience of memory found empirical evidence that the same areas of the brain activated during encoding are also activated during remembering (Buckner & Wheeler, 2001).[24]

Researchers have found that episodic memory is most reliable when an agent uses careful attention during encoding to remember as many details as possible, and when the subject relearns details against externally verified means (e.g. listening to a recording of a concert in the days after the event). Episodic memory is least reliable when little attention is placed during encoding and the event is not rehearsed subsequent to its occurrence against objective criteria.

---

[23] Aristotle surmised that phantasia (faculty of mental imagery or imagination) was active during both perceiving and recollecting (Sorabji, 2004). William James (1890) argued that remembering involved the reactivation of the same motor and sensory areas of the brain that were triggered in the originating event.

[24] Converging evidence indicate that episodic memory retrieval activates brain regions involved in the processing of sensory and emotional experiences (Danker & Anderson, 2010). Evidence to support the reactivation of these regions comes from several methodologies, including studies of brain-damaged patients, fMRI experiments, and cognitive behavioral studies. However, research on brain-damaged patients is particularly illuminating as it shows that impaired perception leads to impaired episodic memory (Greenberg & Rubin, 2003). Retrograde amnesia in a single modality (e.g., auditory agnosia—the inability to recognize auditory stimuli) can be attributed to impairments in auditory memory. If the deficit is visual (visual agnosia), the memory deficit can be more global, suggesting that autobiographical memory is particularly influenced by the quality of mental imagery (Rubin & Greenberg, 1998). However, it is important to recognize the limitations of neuropsychological studies. Deficits found in particular patients do not allow researchers to precisely identify the cause of the problem. For example, it may be that aphasic patients are able to retrieve relevant sensory components, but cannot *process* or *attend* to those components.

Because episodic memory is essentially a constructive faculty used with inconsistent verification, memories are developed, altered and elaborated upon to accomplish a variety of goals that may or may not align with remembering the truth about the past.

The point of this brief examination of memory processes is to illustrate the great range of cognitive processes that are relevant to evaluating any reliabilist account. What justifies an implicit association is quite different from what justifies an explicit recollection. For example, implicit processes such as association are more akin to the thermometer analogy of basic reliabilism. However, explicit, reflective memory needs a reliabilist account that acknowledges inferential and intellectual higher-order knowledge-producing processes and this can be found in virtue epistemology.

1.6   VIRTUE EPISTEMOLOGY

Virtue epistemology is a variant of reliabilism in which the cognitive circumstances and abilities of an agent play a justificatory role.[25] In sympathy with rationalists, virtue epistemologists argue that there is a significant epistemic project to identify intellectual virtues that confer justification upon a true belief to make it knowledge. However, virtue epistemologists are not opposed to empirical pressure on their theories. The openness of virtue epistemology to developments in cognitive psychology (and cognitive science more generally) encourages the cooperative naturalism advocated in §1.3.

---

[25] Contrast virtue epistemology with pure reliabilism or evidentialism where justification does not depend on agency.

Virtue epistemology aims to identify the attributes of agents that justify knowledge claims. Like other traditional epistemologies, virtue epistemology cites normative standards that must be met in order for an agent's doxastic state to count as knowledge, the most important of which is truth. Other standards include reliability, motivation or credibility. Of the many varieties of virtue epistemology (e.g. Greco, 2010; Zagzebski, 1996) I focus particularly on Ernie Sosa's (2007, 2009a, 2011) account. Unlike other theorists, Sosa specifies a truly iterative account of reliabilist justification by dividing it into two levels: animal and reflective.

1.6.1 ANIMAL KNOWLEDGE

Animal knowledge is based on an animal's capacity to survive and thrive in the environment regardless of higher order beliefs about survival, without any reflection or understanding (Sosa, 2007). An agent has animal knowledge if their beliefs are accurate, they have skill (i.e., are adroit) at producing accurate beliefs, and their beliefs are apt (i.e., accurate due to adroit processes). Knowledge is limited to the agent's domains of competence. Animal beliefs constitutive of knowledge are functional—and retain teleological aims such as truth—however, they involve no free action by the agent.

An obvious interpretation of animal knowledge might be to suppose that Sosa is referring to implicit memory processes, particularly associationist or procedural knowledge-how. However, because Sosa always discusses his theories with reference to *belief*, I interpret his discussion of competencies as metaphors or analogies intended to motivate the treatment of doxastic states via reliabilist

means, not literal statements about how knowledge is represented. Throughout this research I treat animal knowledge as implicit, representational beliefs (see, §1.5.1.4) characteristic in humans and perhaps the higher animals[26].

1.6.2 REFLECTIVE KNOWLEDGE

Reflective knowledge is animal knowledge plus an "understanding of its place in a wider whole that includes one's belief and knowledge of it and how these come about" (Kornblith, 2009, p. 128). Reflective beliefs constitutive of knowledge are functional and judgmental. Reflective knowledge draws on internalist ideas about justification (e.g. intuition, intellect and so on) in order to bolster and improve the epistemic status brought via animal knowledge alone. A reflective belief does more epistemic work than an animal belief "if it coheres properly with the believer's understanding of why it is true (and, for that matter, apt, or true because competent), and of how in which it is sustained as reliably truth-conducive" (Sosa, 2009a, p. 138). Reflective knowledge encompasses all higher order thinking (metacognition), including episodic memory, reflective inference, abstract ideas, and counterfactual reasoning.

To satisfy naturalistic reliabilists, Sosa's virtue epistemology depends on empirically plausible decision-making processes. Indeed, animal and reflective knowledge comport with two distinct decision-making systems: (mostly) implicit System 1, and explicit System 2 (Evans & Frankish, 2009; Kahneman, 2011; Stanovich, 1999; Stanovich & West, 2000).

---

[26] Perhaps some animal knowledge, like background beliefs, be brought to explicit consideration with attention and reflective processes?

System 1. Operates automatically and quickly, with little or no effort and no sense of voluntary control.

System 2. Allocates attention to the effortful mental activities that demand it, including complex computations. The operations of system 2 are often associated with the subjective experience of agency, choice, and concentration

System 1 operates in the background of daily life, going hand-in-hand with animal knowledge. Kahneman offers the following list of System 1 capacities (Kahneman, 2011, p. 21):

> Detect that one object is more distant than another
> Orient to the source of a sudden sound
> Complete the phrase "bread and…"
> Make a 'disgust face" when shown a horrible picture
> Detect hostility in a voice
> Answer to 2 + 2 = ?
> Read words on large billboards
> Drive a car on an empty road
> Find a strong move in chess (if you are a chess master)
> Understand simple sentences
> Recognize that a "meek and tidy soul with a passion for detail" resembles an occupational stereotype.

These implicit decisions draw on a range of memory systems. Some are non-representational (e.g. orienting to a sudden sound, driving a car on an empty road), while others involve background beliefs (e.g. completing the phrase 'bread and ....'). Though we are born "prepared to perceive the world around us, recognize objects, orient attention, avoid losses, and fear spiders" (Kahneman, 2011, p. 21), System 1 grows as we grow. It learns associations, skills and short-cuts to improve our efficiencies and as we become experts (e.g., at chess), System 1 is capable of tasks that take a normal person a lot longer to figure out. System 1 accesses a vast range of information about our culture, our beliefs, and our general knowledge automatically and without effort. Many of System 1's actions are involuntarily performed, such as reading a billboard or knowing that 2 + 2 = 4.

Other activities (such as chewing) are normally involuntary, but they can be brought under voluntary control. Despite these differences, both System 1 and System 2 involve attention. Orientating oneself to a loud sound is the reaction of System 1 and it is the efforts of System 2 that return us back to our conversation. All of System 2's work is done during conscious attention.

System 2 requires our attention and it is disrupted when attention is drawn away. System 2—the conscious reasoning self—makes choices, decides what to think about and what to do and it can construct ordered thoughts (such as a sequence of steps). System 2 receives input and suggestions from System 1, but it sometimes overrules these suggestions. Here are Kahneman offers the following examples of System 2 operations:

> Brace for the starter gun in a race
> Focus attention on the clowns in the circus
> Focus on the voice of a particular person in a crowded and noisy room
> Look for a woman with white hair
> Search memory to identify a surprising sound
> Maintain a faster walking speed than is natural for you
> Monitor the appropriateness of your behavior in a social situation
> Count the occurrences of the letter a in a page of text
> Tell someone your phone number
> Park in a narrow space (for most people except garage attendants)
> Compare two washing machines for overall value
> Fill out a tax form
> Check the validity of a complex logical argument

In each of these cases, System 2 is engaged and conscious effort is required from the agent. System 2 helps us to do things that 'do not come naturally.' And resolving System 2-type problems require continuous exertion and attention. As a result, there are strict limits on how much a person can do with System 2 at any one point in time because many effortful activities interfere with one another. As Kahneman points out, it is hard to calculate the product of 17 x 24 while making a

left turn into dense traffic, but, we can have a conversation with a passenger whilst driving on an empty highway.

Current theories of decision-making suggest that humans do have a division between (mostly) implicit animal processes and explicit reflective processes. As discussed in §1.4, a significant issue for reliabilists is a lack of sophistication about how cognitive processes operate, but reliabilism need not be simple. Sosa's virtue epistemology is not simple. He ambitiously seeks to reconcile the appeal of rationalist intuitions with mechanistic, biological processes. Such a reconciliation would offer a way of explaining how higher order thoughts and reflections can impart justificatory power upon doxastic states satisfying to a reliabilist empiricist. To a large extent, both the interest in and the criticisms of Sosa's account arise from evaluating both the value and appropriateness of this 'higher order' reliabilism.

There are two types of criticism leveled at Sosa's virtue epistemology 1) Is aptness sufficient for justification? And 2) Is aptness sufficient for knowledge?—and critics wonder if Sosa incorporates all epistemic theories to the benefit of none (Kornblith, 2009)? One of the answers this dissertation seeks to ascertain whether justificatory tools from Bayesian epistemology (and decision-theory) can supplement the justification of virtue epistemology. Sosa hopes that aptly apt belief can do some philosophical heavy lifting; it may be that Bayesian methods can supplement the hydraulics.

I delve into decision-theory because virtue epistemology is focused on— perhaps even defined by—justification at the moment of decision. Beliefs are

justified by competencies expressed by an agent in a particular performance and for an action to be justified, it must be apt. Being apt means that the action is accurate due to adroit processe, and adroitness is best understood through the lens of decision-theory.

1.7 BAYESIAN EPISTEMOLOGY

1.7.1 MOTIVATION

Bayesian epistemology argues that typical beliefs exist (and are performed) in degrees, rather than absolutes, represented as credence functions (Christensen, 2010; Dunn, 2010; Friedman, 2011; Joyce, 2010). A credence function assigns a real number between 0 and 1 (inclusive) to every proposition in a set. Thomas Bayes argued that our success in the world depends on how well credence functions represented in our minds match the statistical likelihoods in the world. He famously said:

> If a person has an expectation depending on the happening of an event, the probability of the event is to the probability of its failure as his loss if it fails to his gain if it happens (Bayes, 1763, p. 4).

By aligning ourselves with the uncertainties in the world, we become powerful agents to direct ourselves and others. The ideal degree of confidence a subject has in a proposition is the degree that is appropriate given the evidence and situation the subject is in. Of course, no one is an ideally rational agent, capable of truly representing reality, so our goal instead, is to revise and update our internal representations in response to confirming and disconfirming evidence, forging ahead towards ever more faithful reconstructions of reality.

Bayesian epistemology encourages a meek approach with regards to evidence and credences. As Hajek and Hartmann (2009) argue, "to rule out (probabilistically speaking) a priori some genuine logical possibility would be to pretend that one's evidence was stronger than it really was." Credences have value to an agent, even if they are considerably less than 1, and therefore are not spurned. Contrast this with the typical skeptic in traditional epistemology whose hunches, suppositions and worries can accelerate the demise of a theory of knowledge, regardless of their likelihood.

Bayesian epistemology has several advantages over traditional epistemology in terms of its applicability to actual decision-making. Firstly, Bayesian epistemology incorporates decision-theory, which uses subjective probabilities to guide rational action and (like virtue epistemology) takes account of both our desires and our opinions to dictate what we should do. Traditional epistemology, meanwhile, offers no decision theory, only parameters by which to judge final results. Secondly, Bayesian epistemology accommodates fine-grained mental states, rather than binaries of belief or knowledge. Finally, observations of the world rarely deliver certainties and each experience of the world contributes to a graduated revision of beliefs. While traditional epistemology requires an unforgiving standard for doxastic states, Bayesian epistemology allows beliefs with low credences to play an evidential role in evaluating theories and evidence. In sum, the usefulness of Bayesian epistemology lies in its capacity to accommodate decision-theory, fine-grained mental states and uncertain observations of the world.

A comprehensive epistemology will not merely specify the conditions in which beliefs are justified; it will also offer normative guidance for making rational decisions. Bayesian epistemology (incorporating both confirmation theory and decision theory) is a comprehensive theory of decision-making that links beliefs to the best course of action, and it is the 'only theory in town' to do so. As such, any epistemology wishing to be comprehensive ought to hitch itself to the Bayesian wagon.

1.7.2 BAYES' FORMULA AND CONDITIONALIZATION

While there are many kinds of Bayesian, and many distinctions between them, they all depend on Bayes' formula. Additionally, they depend on Kolmogorov's axiomatization (for more detail see Hajek & Hartmann, 2009, p. 94). Kolmogorov defines the *conditional probability of a given b* by the ratio of unconditional probabilities:

$$P(a|b) = \frac{(P(a \cap b)}{P(b)}, \text{ provided P(b)>0}$$

If P(a|b) = P(a), then a and b are said to be independent (relative to P). For example, if the probability of it raining today, given that I am wearing a blue shirt is the same as the probability of it raining today regardless of the color of my shirt, then these likelihoods have no causal interactions with each other. Versions of Bayes' theorem can now be proven:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

$$P(a|b) = \frac{P(b|a)P(a)}{P(b|a)P(a) + P(b|\sim a)}$$

If there are many hypotheses being considered (h_1, h_2,…,h_n), and evidence e then for each j

$$P(h_i|e) = \frac{P(e|h_i)P(h_i)}{\sum_{j=1}^{n} P(e|h_j)P(h_j)}$$

The $P(e|h_i)$ terms are called likelihoods, and the $P(h_i)$ terms are called priors (Kim, 1988). Bayesianism describes the operation between a piece of evidence e and the confirmation of a theory or hypothesis h as:

e confirms h (relative to P) iff $P(h|e) > P(h)$.

That is, a piece of evidence confirms a hypothesis if and only if the probability of the hypothesis given the evidence is greater than the probability of the hypothesis in the absence of evidence (regardless of whether the evidence is absent or non-existent).

Bayesianism offers a way of interpreting an agent's degrees of belief of an agent at a given point in time t. However, it also offers a process that explains how these degrees of belief change over time. The synchronic thesis claims that beliefs are probabilities, while the diachronic thesis explains how these probabilities are updated via conditionalization.

Conditionalization means that an agent's posterior subjective probability—after taking account of evidence X, P*new*—is to be set equal to her prior conditional probability P*old*(·|X) (Greaves & Wallace, 2006). For example, my

subjective probability P*new*—that it will rain today *after* watching the weather report X—ought to equal my conditional probability that it will rain today given that weather report X. Conditionalization tells us how to operate on evidence (i.e. revise our beliefs), whatever likelihood the evidence has. It offers a rational way to alter degrees of belief with uncertain evidence. Bayes' rule works with conditionalization to operationalize how any piece of evidence should affect our beliefs, based on how likely the outcome is, both with and without the evidence. For example, if the weather report is notoriously unreliable, then the impact of such a report on my belief that it will rain ought to be less than if the weather report was highly reliable. This is true regardless of what the weather report predicts.

Bayes' rule provides a way of taking multiple lines of evidence into consideration when revising beliefs, such as one's own perception of the weather conditions (clouds in the sky outside one's window, the force of the wind against one's skin), one's general knowledge of rain patterns in the area given the time of year, or the testimony of others. With these tools, Bayesian epistemology gives a justification of ideal scientific practice (i.e., solving the problem of induction via confirmation theory), and of decision-making. It also helps to resolve issues including how to quantify the value of coherence between beliefs—which I examine in *Chapter 3*.

Bayesian epistemology endeavors to find ideally rational, formal methods by which to make decisions using a probability calculus. Confirmation theory offers an abstract conception of an ideally rational agent, who reasons in a

probabilistically consistent way. A Bayesian agent treats each new problem with the same probabilistic apparatus, and when outcomes are uncertain, the agent ought to rank the likelihoods in the same way, given the same evidence.

An agent who violates the laws of probability will select actions that guarantee a loss—what is called the *Dutch Book argument*.[27] The basic idea of a Dutch book is to imagine a bookmaker encountering an irrational punter. If the punter's betting behaviour violates the axioms of probability, then the bookmaker can guarantee a profit over the long term. Inconsistently attributing likelihoods makes a person irrational and also ensures they fail to meet their goals. There are rational means to estimate likelihoods in the absence of evidence. For example, *the principle of indifference* asserts that if you have a poverty of information and several options to consider, then you assign each option equal probability until subsequent evidence gives you reason to differentiate between them.

In its purest form, Bayesian epistemology sets a very high standard for humans to meet in order to be ideal, rational agents, and, as a result, it is considered by some to be an inappropriate model for epistemology to adopt (Goldman, 1978). Humans consistently deviate from standards set by a rationality based on self-interest. For example, economic organization and the structure of social interactions play a sizable role in individual priorities and decisions. For example, humans exhibit greater prosocial behaviors when markets are highly integrated and cooperation yields high pay-offs (Henrich et al., 2005). As such, it may be that the complexities of real individuals in diverse circumstances mean

---

[27] The dutch book argument stems from Ramsey's (1964) work in "Truth and Probability".

that 'pure' Bayesian epistemology may not fully inform an agent-centered account of epistemic justification.

Putting aside the advantages and limitations of the normative Bayesian project, does human cognition even approximate Bayesian processes? What evidence exists that our mind is Bayesian? It is important to distinguish here between *computational*, *algorithmic* and *instantiation* levels of cognition (Fodor & Pylyshyn, 1988). The *computational level* refers to what a system does (what problems cognition solves or overcomes and why). The *algorithmic* (or *representational*) *level* refers to how a system solves these problems, that is, what processes cognition employs to solve problems. The *instantiation* (or *physical*) *level* refers to how a system performs these calculations within its physical make-up (e.g., in the case of vision, what neural structures or activities implement the visual system). In the paragraphs below, I look at each of these levels in turn.

1.7.3 COMPUTATIONAL LEVEL

On a computational level, the brain must represent probabilities to process information (Clark, 2013; Gallistel, 2012; Gallistel & King, 2009).[28] That is, the brain must somehow represent the set of possible messages and a probability distribution over that set, in order to make decisions. Consider a hunter listening in the woods to the sounds of cracking sticks and crunching leaves. The hunter will inductively (and implicitly) assign likelihoods to these sounds based on prior experiences in the woods. Given the volume of sound, the pattern of silences and the speed the sound is travelling, the hunter assigns likelihoods that it is a raccoon

---

[28] According to Shannon's (1948) theory of communication.

(P=0.4), a deer (P=0.2) or a squirrel (P=0.4) for a total probability = 1. If the brain has no expectation of the messages that *might* be sent, it is unable to disentangle information from noise. Even a novice hunter with no experience will make assumptions about the source of the noise (e.g., perhaps they assign a high likelihood that it is a bear?).

An informative signal changes the receiver's probability distribution about the possible states of the world. For example, a hunter who hears something scratching up a tree, might reduce the options to a raccoon (P=0.5) and a squirrel (P=0.5). An uninformative signal offers no change in the receiver's probability distribution. The novice hunter might be too frightened to process the arboreal sounds, still fixating on the terrifying possibility of a bear, waiting in the undergrowth.

The most *informative* signals are those that reduce the possible set of messages the most dramatically. This is a Bayesian operation—the receiver's representation after an informative signal is their posterior probability distribution over the possible values of an empirical variable (such as animals in a forest). Crucially, the amount of information communicated is the difference in uncertainty between the prior probability (before the signal was sent) and the posterior probability (after the signal was sent). For example, hearing arboreal scratching is less informative than seeing a stripy tail, because seeing the latter reduces the options the most, i.e. a raccoon (P=1). The likelihood is the probability of getting a particular signal given the state of the world. The

unconditional probability is the overall probability of getting a particular signal regardless of the state of the world.

Accepting that the brain must incorporate Bayesian principles to some degree, the sufficiency of Bayesian psychology remains an issue. Humans are flexible thinkers—able to change their mind, go back on previous beliefs and be influenced by other beliefs. Classical conditionalization, meanwhile, is strict and irreversible—"no learning episode can be undone by any subsequent application of the classical rule" (Döring, 1999, p. S380). Human psychology has two features that conflict with classic conditionalization 1) *confirmational holism*: background beliefs influence the justification of occurrent beliefs, and 2) *commutativity:* the order in which information is learned ought not affect the outcome. On the subject of commutativity, Jonathan Weisberg (2009) points out that "it shouldn't matter whether I find the murder weapon in the maid's room first and then hear testimony about her alibi, or the other way around. Either way my ultimate attitude about her guilt should be one of guarded suspicion" (p.794).

An 'unlearning' modification to conditionalization was suggested by Jeffrey (1983)—*Jeffrey conditionalization*—and aims to ensure conformational holism with repeated applications of the same rule of conditionalization. However, Frank Döring (1999) argues that *Jeffrey conditionalisation* fails to maintain *commutativity*, because it gives different probability distributions depending on the order of events. Additionally, Weisberg (2009). argues that the rigidity of conditionalization (both Jeffrey and standard) renders it inherently anti-holistic.

Many authors (Hawthorne, 2004b; Wagner, 2013; Zhao & Osherson, 2010) have defended Jeffrey conditionalization against these (and other) attacks. Wagner, in particular, argues that Weisberg has been a victim of parochialism—a sin that Jeffrey (1983) explicitly warned against. Parochialism occurs when evidence relevant to another event is ignored due to a poor choice of partition. Weisberg's choice of partition makes *Jeffrey conditionalization* appear anti-holistic, but, Jeffrey acknowledged that he had no procedure for picking the relevant partition—an example of the frame problem.[29] Note that adopting *chorism*—the view that a *limited* set of evidence ought to be sufficient for justification (see §1.2)—is an alternative to holism that retains the usefulness of sets of influential beliefs.

The 'frame problem' arose in early applications of logical methods in artificial intelligence (McCarthy & Hayes, 1969). Consider an agent, with a set of beliefs S, who makes an action, A (e.g. turns a cup upside down). What other information in the set of beliefs needs to be updated in order to take account of this action? Almost all other knowledge seems irrelevant and should remain unchanged (e.g., that the street is empty or that the burglar alarm is off). Logically, however, it is possible that A has many consequences. For example, if the cup is valuable and located in an alarmed glass case, then turning it may set off the burglar alarm and draw crowds of interested bystanders to the street. The frame problem, therefore, describes the difficulty of creating a procedure that can distinguish relevant from irrelevant connections between beliefs to inform action.

---

[29] Even if holism can be resolved, Jeffrey still faces the conundrum of *commutativity*.

Fodor (1983) added that incorporating a new belief can overthrow pre-existing beliefs in almost any dimension, and in ways that are impossible to predict. The frame problem is arguably the most intractable problem facing any theory of cognition, and although recent advances by search engines have shown breakthroughs in information search (Shanahan, 2009), little headway has been made with regards to information choice in humans (and animals)[30].

Another problem for the Bayesian approach is the tendency of researchers to treat all evidence equally, combining *any* available data with prior beliefs (including those about causal structure) (Bes, Sloman, Lucas, & Raufaste, 2012). It may be that judgments are heavily influenced by beliefs about causal structure and only then are they influenced by statistical data (Pearl, 2000; Suppes, 1970; Wheeler & Scheines, 2013). Agents use a 'causality heuristic,' which means that they only pay attention to stochastic information when it alters their causal beliefs. Causal knowledge influences counterfactual reasoning and interventions, both of which are crucial components of reflective knowledge. In *Chapter 2* I argue that hierarchical Bayesian models that incorporate causal structures and statistical information are a tractable computational account of reflective knowledge.

1.7.4 ALGORITHMIC LEVEL

Whatever the computational arguments, decades of research suggest that Bayesian standards are unintuitive, particularly with regards to base rate information (Kahneman, 2011; Kahneman, Slovic, & Tversky, 1982; Kahneman

---

[30] See (Xu & Wang, 2012) for a recent attempt to deal with the frame problem using a Non-Axiomatic Reasoning System (NARS)

& Tversky, 1996). Base rates are the likelihood of a phenomenon, without regards to properties of an individual. For example, the base rate of clear, sunny days in Perth, Western Australia, is 131 per year, regardless of the weather on any particular day. Knowing the base rate of clear, sunny days in Perth helps us to predict the likelihood of clear, sunny days on any arbitrary day in the future. Base rates matter, but unfortunately in ways that are not immediately obvious. Consider the following:

> You see a person reading the Economist on the New York subway. Which of the following is a better bet about the reading stranger?
>> She has a PhD
>> She does not have a college degree

Many people would pick the PhD option because they tend to guess based on stereotypical rather than statistical information. In fact, it is statistically far more likely that the commuter reading the magazine has no college degree. Because the base rate of commuters without college degrees is far higher than the base rate of commuters with a college degree. People make this mistake because base rates are not as salient as the narrative of an individual. People typically overestimate the value of their evidence (e.g., reading the Economist) and underestimate the pull of associative coherence (e.g., the desire to make a compelling story), particularly for unlikely (low base-rate) events. Even Daniel Kahneman (2011) relates to this problem, explaining that "even now I find it unnatural [to consider base rates and question the diagnosticity of evidence]" (p. 154). People prefer to explain events with causal stories rather than statistically appropriate inference (p. 166-170). The depressing difficulty of statistical inference is summed up by Nisbett and Borgida (1975), who argue that "subjects'

unwillingness to deduce the particular from the general was matched only by their willingness to infer the general from the particular." While humans are enthusiastic to generalize from a single instance, they are reluctant to apply likelihoods in any particular case.

Gerd Gigerenzer's lab (Gigerenzer, 2008; Gigerenzer & Brighton, 2009; Gigerenzer & Hoffrage, 2007; Gigerenzer & Sturm, 2012) has been the most vocal opponent of the view that Tversky & Kahneman-style (1982) experiments show that humans are largely irrational, arguing that the design and assumptions of these experiments are inherently flawed. For example, Gigerenzer showed that people can reason according to Bayesian norms when base rates are more visible and when discussions are framed in frequentist terms (Gigerenzer, 1991).[31] Note that Gigerenzer does not dispute the normative power of Bayesian rationality in this case, merely the ecological validity of the tests.

Ecological validity is a cornerstone of bounded rationality; the notion that an agent's success ought to be evaluated relative to their capacities, goals and information environment.[32] There is no point having a theory of rationality that depends on information processing far outside the capacities of the agent because the typical information environment humans face is far removed from Laplace's world of superintelligence where all past information must be considered.[33]

---

[31] Bayesianism and frequentism have, in the past, been considered incompatible, but Williamson (2013) argues that they should exist in harmonious symbiosis. Bayesians need frequentist information to calculate prior probabilities, and frequentists need Bayesian methods to calculate the odds of a single case.

[32] See also philosophical discussions of Minimal Rationality (Cherniak, 1981).

[33] "We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it, nothing would be uncertain, the future as well as the past would be present to its eyes. The perfection that the human mind has been able to give to astronomy affords but a feeble outline of such an intelligence" (Laplace, 1820/1951).

Instead, the world presents puzzles that lie somewhere between pure chance and pure strategy. In the actual world, Gigerenzer argues that heuristic processes can (and must) outperform decision-making based on crunching data. In fact, Gigerenzer and Sturm (2012) defend the normative use of heuristics when problems are computationally intractable (e.g., calculating the optimal sequence of chess moves), when circumstances lack robustness (e.g., when there is high uncertainty and a small sample size, such as predicting the stock market with ten years of data), or when problems are ill-defined (e.g., trying to find the best job, optimize a website, or buy the ideal car is hampered by a lack of knowledge about the alternatives, probabilities, consequences and utilities).

All arguments in defense of heuristics stem either from descriptive or normative advantages (Gigerenzer & Sturm, 2012 Table 1). Descriptively, heuristics are defended as more explanatory and therefore more predictive than other models. For example, the *default heuristic* explains why marketing campaigns fail to increase organ donor registration. Normatively, heuristics are defended as outperforming other strategies, meaning that they are either more accurate, or more effective and efficient for the agent. For example, in a summary table of heuristics, Gigerenzer defends: *The recognition heuristic*, *fluency heuristic*, *satisficing*, *imitate the majority* and *imitate the successful* because they are beneficial to the agent in achieving their goals or aims. In the same table:

Take-the-best, *tallying*, *tit-for-tat* and *equality heuristic* are pitched as matching or improving on the accuracy of other models[34].

Reconciling the usefulness of traditional epistemic values, such as accuracy and rationality, has led Gigerenzer and Sturm (2012) to a cooperative naturalized approach. They argue that empirical research ought to inform norms of rationality in some domains—as well as providing guidance for how human reasoning could be improved. However, in other domains, the 'standard' conception of rationality involving logic or probability ought to apply (Stein, 1996).

1.7.5 IMPLEMENTATION

To meet computational and algorithmic demands, the brain needs both Bayesian processes and symbolic representation—a read/write memory that can retain information and retrieve it for a later time (Gallistel & King, 2009). Due to the speed and complexity of human cognition, the physical basis of computation must achieve:

> high information density, (bits per cubic micron). It must have low volatility; hence, a low use of energy to maintain the physical changes that carry the information forward in time. It must be capable of entering either directly or via transcription into physically realized compact procedures for implementing two-argument functions (p.286).

Gallistel and King are skeptical that these parameters are met by neural connections or connectionist architecture. A better model for how information is biologically processed is found in genetic code (p.125), because genetic mechanisms hold heritable information and direct the growth and development of living things by reading, transcribing and translating code into protein structures.

---

[34] Models include: Optimization models, multiple regression, neural networks, exemplar models and decision tree algorithms.

The scope, limits and demands of Bayesian cognition are relevant to this dissertation for a number of reasons. Rationality is ecologically constrained, but it is tested against traditional epistemic aims such as truth and reliability. Reliable beliefs are produced algorithmically by many processes including Bayesian, heuristic, inferential, and associationist processes, and so forth. Bayesian epistemology makes a substantial contribution to the normative standards for decision-making at a computational level, regardless of what occurs on either the algorithmic or biological instantiation and the justificatory status of beliefs is partly determined by the degree to which their production abides by the dictates of logic and probability theory.

Reliability can do a lot of work to justify beliefs, but reliability alone is not enough. Calibrating traditionally internalist values of reflective knowledge (e.g. logical consistency) and non-arbitrary attribution of prior probabilities remains an issue for a hybrid Bayesian Virtue epistemology, and I try and resolve this weakness in *Chapter 3* where I examine how virtue and Bayesian epistemologies bolster reliabilism with coherence.

1.8   CONCLUSION

I started this chapter with the notion that human flourishing depends on knowledge, but that *certain* knowledge may be unobtainable because humans flourish in informationally impoverished environments. In this dissertation, I examine our understanding of knowledge and our understanding of justified epistemic states that fall short of knowledge. In particular I focus on a virtue-

Bayesian epistemology—the justification of animal and reflective knowledge through iterative Bayesian processes.

While a Bayesian virtue epistemology seems promising, it faces some concerns: Is apt belief apt enough for knowledge?—addressed in *Chapter 2*; how can coherence add justification to a reliabilist account?—addressed in *Chapter 3*. What emerges from these investigations is homeostatic epistemology. Homeostatic epistemology argues that apt belief *is* apt enough for knowledge, Epistemic justification arises from the reliable coordination of limited sets of cohering, foundational beliefs, and it is the ebb and flow of coherence and distinctness between these beliefs that leads to homeostatic epistemology, which I articulate in *Chapter 4*.

2   BAYESIAN VIRTUE EPISTEMOLOGY

Hornblower took his turn at the sheet, but he would not trust himself with the tiller, especially when night fell; he knew he had not the knack of keeping the boat on her course by the feel of the wind on his cheek and the tiller in his hand (Forester, 1950)

2.1   INTRODUCTION

Reflective knowledge is the pinnacle of human functioning, and it is traditionally conceived of as the reservoir of the a priori—a revered, almost mystical mental faculty through which Platonic ideals, truths, and axioms depart the heavens and settle on the brow of mortal man. In contrast, reliabilist beliefs are merely dumb associations, forged by mechanistic repetition of limited cognition, that create impoverished models of the external world. In this chapter, I splice together these apparently conflicting processes by examining Sosa's reliabilist account of reflective knowledge. To do so, I draw on another agent-centred, normative and reliabilist epistemology—Bayesian epistemology.

Typically, the study of knowledge as 'justified true belief' has assumed that beliefs are all or nothing, rather than probabilistic (BonJour, 1985; Conee & Feldman, 2004; Goldman, 1986). The view that knowledge must be certain stems from Socrates, but it was Descartes who defined modern analytic philosophy through his efforts to seek certainty amidst powerful skeptical objections (Hajek & Hartmann, 2009)—the all-or-nothing view. However, Descartes had competition to set the agenda. Within four years of his death, Pascal and Fermat argued that one should strive for rational decision-making, rather than truth—the probabilistic view.

The all-or-nothing and probabilistic views are illustrated by the difference between Descartes' Cogito and Pascal's wager. In the Cogito, Descartes argues for a belief in God based on rational reflection and deductive reasoning. In the wager, Pascal argues for a belief in God based on outcomes evaluated probabilistically. Sosa's virtue epistemology applauds the Cartesian quest for knowledge, but it does so with a competence-based epistemology.

Sosa divides knowledge into categories of animal and reflective belief. More radically, he suggests that what justifies animal belief also justifies reflective belief, albeit on a higher level, and he hopes that this distinction can resolve a number of issues that arise from other epistemic accounts. As a reliabilist, he avoids the inductive and deductive limitations of a completely internalist account of justification. By valuing the distinct contribution of the intellect, he avoids the simplistic outcomes of brute associations. Like decision-theorists, Sosa argues that each *performance* of belief ought to be evaluated with regards to the agent experiencing it. Sosa believes that a competence-based virtue epistemology can circumvent difficulties faced by traditional accounts, including a way to explain the absence of knowledge in Gettier-cases.

Bayesian epistemology offers a framework for rational-decision making that dovetails with a competence-based virtue epistemology. Epistemology becomes not just the study of justified true belief, but also the study of the processes of belief revision in response to confirming or disconfirming evidence. Justification arises from the apt performance of reliable processes and their coherence with other beliefs.

This chapter examines the success of Sosa's justification of animal and reflective knowledge. Critics of his view argue that animal beliefs are too weak to do the work of knowledge, or that reflective knowledge represents nothing more than iterations of the same defective stuff. I respond to these criticisms by suggesting that Bayesian processes could supplement perceived deficiencies in the virtue epistemology model. From this, a hybrid Bayesian virtue epistemology emerges.

In §2.2, I provide some background information on knowledge considered as justified true belief. In §2.3, I give more detail of animal knowledge and discuss whether it is sufficient for knowledge. In §2.4, I describe reflective knowledge and address concerns of infinite regress, the problem of luck, the KK principle, and finally the concern that reflective knowledge is just animal knowledge 'twice over.' In §2.5, I defend the 'theory theory' of reflective knowledge. In §2.6, I use a theory theory account of reflective knowledge to defend reflective knowledge against the new evil demon skeptical scenario and propose some normative guidelines. In §2.7, I conclude by outlining the problem of coherence as a source of justification to be addressed in the subsequent chapter.

2.2   BACKGROUND

What is it to know? Prairie dogs are cooperative, social rodents with an astonishing capacity to communicate idiosyncratic details of individuals of different species. Their alarm calls can discriminate the color of clothes worn by a particular human, their shape, their size, and even whether a human has ever appeared with a gun. As an example, when human observers were unable to

distinguish between a German Shepherd and a coyote on a particular occasion, the prairie dogs made the distinction immediately (Slobodchikoff, Perla, & Verdolin, 2009). As this example demonstrates, it seems reasonable to assert that prairie dogs *know* who is walking among them.   When we say this, however, do we really mean that they *know*, or are we anthropomorphizing the prairie dogs' calls and actions to suit our own preference for narrative? After all, I hardly think my cat really *knows* that I'm coming to bed, but I use the term regardless.[35]

Our varied use of the word 'know' should not be confused with its reference. What is it to know? Most people would agree that knowing refers, in some way, to the truth (e.g., if the prairie dog knows who is approaching, they must be accurate). I take this as a given. The second idea is that knowledge is propositional (i.e. that knowledge is the domain of creatures with *beliefs* about the world) and this is a far more controversial view. Gilbert Ryle (1946/1971, 1949) suggested that at least two kinds of knowledge exist: knowledge-that and knowledge-how.[36] Empirical work on memory systems confirms Ryle's original distinction and expands upon it (see §1.5). In this chapter, however, I only consider propositional knowledge (whilst acknowledging that other sorts of knowledge may exist).

Knowledge consists of true beliefs. 'Belief' is a propositional attitude we have when we take something to be true (Schwitzgebel, 2011), and I take beliefs to be both functional and representational. Functionalism about mental states is

---

[35] I daresay that he *anticipates* that I will join him.

[36] There has been philosophical pushback against this division, claiming that knowledge-how ascriptions attribute propositional knowledge (see Stanley, 2011). However, for the sake of my discussions, I will take it as a given that "we know much more about the way the world is than we do about the semantics of our talk about that world" (Devitt, 2011). Although as Ernest Sosa points out (personal communication August 15, 2013), all knowledge may be propositional, even if people fail to adequately ascribe them.

the view that beliefs are causally related to sensory stimulations, behaviour and other mental states (Armstrong, 1968; Fodor, 1968b). Representationalism is the view that beliefs represent how things stand in the world (Fodor, 1975; Millikan, 1984). This means that a true belief that there is a coffee cup is connected to the actual fact that there is a coffee cup. A belief that there is a coffee cup may cause an agent's desire for coffee, prompting his hand to move to pick it up and the feeling of cold porcelain against his skin. In this chapter, I remain agnostic on the specifics of what particular variant of functionalism or representationalism ought to be adopted, as neither debate is central to the arguments of this chapter. For the purposes of this chapter, the important ideas are twofold (1) that beliefs are causally connected to our experiences and (2) that beliefs are internal representations of the world.

Beliefs are also propositional—either language-like or map-like. Like language, beliefs can be combined in an infinite number of arrangements (productivity) and exhibit logical structure (systematicity) (Fodor, 1975; Fodor & Lepore, 1996). Beliefs also seem to be map-like, in that changes to one belief alter its relationships with other beliefs (just like moving a checkpoint on a hike simultaneously changes its relationships with all other checkpoints and features on a map) (Bradden-Mitchell & Jackson, 1996).

Tying knowledge to propositional, functional, and representational belief means that animals such as sea slugs, insects, or bacteria do not have knowledge, although they can react to their environments in remarkably complex ways inviting one to use terms like 'believe' and 'know.' On the other hand, higher

order animals such as chimpanzees, dolphins, and perhaps even prairie dogs do have beliefs.[37] Most importantly, young children at the age of three have beliefs, even though they have no concept of belief and their broader understanding of the world is deficient (Marcus, 1995). If we grant prairie dogs and children the capacity for belief and the capacity to believe accurately, the possibility exists that they can, in fact, know.

Nonetheless, true belief is not usually considered sufficient for knowledge. Suppose my friend flips a coin, hides the result under his hand, and asks me whether I believe the coin came up heads or tails. I may *believe* 'heads' and, indeed, perhaps the coin *is* heads, but I do not *know* that it is heads until my friend lifts his hand and I can perceive Washington's profile. Epistemologists are particularly interested in justification—i.e., what has to be added to true belief in order to create knowledge.

Perception, memory, rational processes and testimony are all sources of justification. In the coin example, what justifies my belief that the coin landed heads up is the reliable operation of my perceptual capacities in normal lighting conditions. Epistemologists argue about whether perceptual knowledge is different to higher order reflective knowledge—do my low-grade beliefs about the existence of my hand require the same sort of justification as my lofty high-grade beliefs about the existence of objective chances? Typically, epistemologists use different accounts of justification depending on the sort of knowledge they are trying to explain, because it is difficult to explain how low level associations

---

[37] Though it is plausible that prairie dogs have beliefs *de re*, but not *de dicto* (Davidson, 1984).

could result in justified rational thought and, conversely, how abstract thinking could explain perceptual knowledge. Much of this chapter will explore justification for two kinds of knowledge: animal and reflective.

Before I articulate the details of animal knowledge and reflective knowledge in the sections that follow, it is important to note here that I assume that knowledge must be considered from the perspective of the agent who holds the belief, rather than as an isolated, objective fact. Knowledge is attributable in the same way that a shade of blue B is understood.

> B = a subjective visual experience within an animal A, at a time T, viewing an object O in conditions C.

There is no understanding of the blueness of the sky without reference to the agent experiencing it. Similarly, knowledge (K) is analyzed with reference to the agent holding the belief:

> K = a justified true belief within an animal A, at a time T, in conditions C.

Just because knowledge must be understood with reference to the agent does not make it *subjective*. There are objective facts about the processes and outcomes of how an agent and object interact in the world, and there are objective facts about the attainment of knowledge.

2.3   ANIMAL KNOWLEDGE

Sosa's virtue epistemology begins with his account of animal knowledge. Animal knowledge is a reliabilist solution based on an animal's capacity to survive and thrive in the environment, regardless of higher order beliefs about survival. Sosa asks us to consider an archer shooting an arrow at a target. A shot is apt when it is accurate not because of luck or a fortuitous wind that pushes the

arrow to the center, but because of the competence exhibited by the archer. Sosa takes beliefs to be long-sustained performances exhibiting a combination of accuracy, adroitness, and aptness. Apt beliefs are accurate (true), adroit (produced by skillful processes) and are accurate *because* they are adroit. Aptness is a measure of performance success, and accidental beliefs are therefore not apt, even if the individual who holds those beliefs is adroit. Take, for example, a skilled archer who hits the bullseye due to a gust of wind rather than the precision of his shot, or a nervous student who randomly picks the right answer to a multiple choice question, even though she might easily explain the content in normal conditions. Animal knowledge involves no reflection or understanding. Although animal knowledge can become reflective if the appropriate reflective stance targets it. For example, a person might have animal knowledge that a deer is 300 feet away in the forest, and when questioned, they reflect on their belief and form a reflective judgment that the deer is 300 feet away with the addition of explicit considerations of the distance between trees, lighting conditions and prior experience.

Sosa argues that unreflective knowledge counts for many implicit beliefs in an adult's epistemology, and even explicitly considered beliefs are pieced together intuitively, rather than with formal reasoning apparatus (weighed up against scientific norms, or logical axioms). For this reason, we are unable to verbalize much of our "non-trivial" knowledge, because we lack the precision of vocabulary or concepts to describe what we know to be true. Sosa (2009a) provides an excellent example of this:

> …experienced sailors embody practical lore that they cannot articulate… a certain gestalt look of the environment will prompt practically appropriate inferences: that a storm is coming, say, even if the knowledge embodied must remain inarticulate: the knowledge that when the sea and the heavens look a certain way, a storm is likely brewing (p.72)

Animal knowledge might 'remain inarticulate' and yet yield 'practically appropriate inferences' nevertheless. We know the typical heft of a billiard ball and how it might perform on tasks against surfaces with various degrees of friction, such as carpet, floorboards, or bedding (Sosa, 2009a). We use our animal knowledge to guide our arms to move at a velocity, with a particular grace to suit the task at hand, without any capacity to enunciate the parameters of this knowledge (Sosa, 2009a). The capacity to *explain* our knowledge is the domain of reflective knowledge.

2.3.1 APT BELIEF IS NOT APT ENOUGH

One criticism of Sosa's argument is that apt belief is too weak to do much work. Christopher Lepock (2010), for example, argues that apt belief is not sufficiently knowledge-conferring to act as the foundation on which reflective knowledge is built. It is important to distinguish between two claims: 1) that apt belief is strong enough to do foundational epistemic work, and 2) that apt belief is sufficient for knowledge.

Can apt beliefs perform foundational epistemic work? The answer to this will depend on what sort of epistemic objects count as foundational. Foundational states must increase the capacity of an agent to discern the truth. According to this definition, reliable credences, true beliefs and knowledge all count as foundational. Indeed, even a fairly weak belief (i.e. reliably formed credence) can

be useful because no belief must stand on its own; its reliability is bootstrapped by its relationships with other beliefs. A Bayesian account suggests that a number of weak beliefs, drawn together coherently, can be foundational (Lewis, 1946; Olsson & Shogenji, 2004). If aptness refers to competent treatment of belief with regards to one's evidence, then it could also count as a foundational epistemic object—regardless of whether it produced knowledge, unGettier-proof justified true belief, or reliable credences. I grant that Sosa hopes that apt belief is knowledge-conferring, not just producing rational beliefs. However, there seems to be a need for quasi-apt beliefs to account for reliably formed credences that fall short of knowledge.

Lepock claims that ordinary human knowledge has at least an implicit awareness of one's competencies that a merely 'reliable' apt belief does not have. This 'implicit awareness' supposedly allows us to adapt to different circumstances, but, the concept 'implicit awareness' used by Lepock is confusing. Animal cognition involves implicit, cognitively impenetrable beliefs that play a complex functional role within an organism, allowing it to learn, adapt, and remain flexible in response to a changing environment or competence. Consider, for example, the flexibility and detail of the prairie dog when it communicates to its peers that a particular predator is approaching. Apt belief has the sensitivity and flexibility to adapt to different circumstances, a point I come back to in the next chapter on the topic of coordination. Even without implicit awareness, apt belief allows an organism to adapt to different circumstances, and, as such, can be considered ordinary knowledge.

Does Lepock mean that ordinary knowledge involves some sort of consciously accessible, cognitive penetrability? Perhaps my notion of 'ordinary knowledge' and Lepock's notion differ? It seems that the knowledge defined by Lepock as 'ordinary' is actually reflective and although philosophers frequently discuss reflective knowledge, this in itself does not render it 'ordinary', as in 'most common'. To describe it as such would be foolish, just as it would be foolish to suppose that explicit memory systems are the dominant memory systems relied on by an agent because the philosophy of memory has been dominated by discussions of explicit memory[38]. It seems that animals and humans have the confidence and success that indicates they know much that they cannot articulate or access consciously.

2.3.2 SOSA'S TRICKSTER

One of the advantages of the AAA structure of Sosa's virtue epistemology is that he hopes it will explain why Gettier-cases fail to be instances of knowledge. Sosa (2007) gives an example of Jim and the Jokester. Jim has apt perceptual beliefs in normal conditions, so in normal conditions he believes that the surface of a table in a room is red and it is red. But consider the risk to Jim's apt beliefs if there's a Jokester afoot who can change either (or both) the lighting in the room or the color of the surface of the table, or both, much like a kaleidoscope. At one moment, in normal lighting conditions, the surface really is red; and in another, the surface is white and merely looks red. The skeptical question here is this: In the situation where the surface really is red and Jim

---

[38] For examples of remembering and recollection in the analytic tradition see particularly Russell (1921), Ayer(1956), Sorabji (2004), Suddendorf & Moore (2011), Martin & Deutscher (1966), Reid (1785)

perceives it as red, does Jim know that it is red? Isn't Jim's true belief that the surface is red just a lucky accident (i.e., the table could easily have been white with red lighting), and therefore not knowledge? Sosa believes that Jim has animal knowledge that the surface is red because his perceptual faculties are operating in the appropriate performance context, with the right causal connection to yield true beliefs. Animal knowledge exists and being used, even though Jim's reflective knowledge fails to realize that he is in a tricky environment.

Some commentators might still feel that Sosa's use of the term 'know' is misplaced (Lepock, 2010) and I am inclined to agree with Lepock that something is odd here. Jim may not have knowledge if believing "the table is red" equates to the stable property 'redness,' as derived from red paint or red plastic. A belief in a stably red table will not do Jim much good, because we know that the table is not, in fact, stably red, but can be either white or red depending on the whim of our jokester. In fact, the table is a clear laminate with internal lighting.

On the other hand, if believing "the table is red" equates to the perceptual belief that the table *looks* red at time *t*, then Jim does have knowledge. The property of "looking red at time *t*" allows for stable or unstable redness as properties of the table. In the latter case, "Jim knows that the table is red" simply means that the table looks red and is red in the moment he sees it. However, this does not confer stable knowledge of the table's true properties—i.e., a clear table with red lighting. The ordinary sense of 'know' that Lepock seeks, is one where Jim knows more about the world than what it *seems* to be, but also what the world *actually* is. I suspect Sosa is in trouble if he thinks Jim knows that the table is

stably red. Sosa agrees that Jim might only know that the table looks red, although he points that that this is an objective fact that the table looks red (based on Jim's visual system), not just a subjective judgment by Jim[39]. In the end, Jim's knowledge that the table looks red is best explained by the fact that the table is unstably red and Jim's perceptual capacities were able to adroitly detect this redness.

Apart from explaining instances of knowledge in tricky situations, Sosa's trickster can also be used to demonstrate that what we need to know varies depending on our goals. For example, say Jim was a school child (J_{sc}) on an excursion to an art gallery. The goal of J_{sc} is to report his experience of art. J_{sc} knows enough when he reports accurately that the red table made him feel hungry, warm, and frightened. However, suppose Jim was an art critic (J_{ac}). J_{ac} may know that the table made him feel hungry, warm and frightened when it looked red, but this is not enough to write his review. For a review, J_{ac} needs to know what the table is made of, how often it cycles through white and red, whether that cycle is random or not, why the artist chose the medium and apparatus, and so forth. J_{ac} needs to experience all the fluctuations of the table in order to fully evaluate its success as an exhibit. Alternatively, suppose that Jim was a magician's assistant (J_{ma}). J_{ma} needs to have the animal knowledge of the table's redness, an understanding of materials and cycles, and an understanding of the mechanism that switches the lights on, so that he can build an illusion for the magic show. An agent's goals are critical to evaluating their level of

---

[39] Sosa (personal communication, August 15, 2013)

knowledge. If an agent seeks higher levels of knowledge, their beliefs must be evaluated on those terms.

The way a Bayesian reacts when presented with a table that looks red is to form a set of likely hypotheses about what causes the redness, and to attribute priors from background beliefs. The prior probability that the table is red, given it looks red, ought to be high if other factors about its location and circumstances are usual for normal lighting conditions (for example, if the table were outside during the day or inside near a window, or in a police station where colored lights are unusual and white fluorescent lights are cheap). On the other hand, if the table was observed in a photography laboratory, a teenager's bedroom, an art gallery, a circus or a nightclub, then the observer may have a legitimate basis to assign a much higher prior likelihood to the table being white with red lighting. The Bayesian might also have defensible reasons to withhold judgment regarding the table's true color altogether. As these examples demonstrate, the scope of hypotheses under consideration is hugely impacted by the context of experiences. The Bayesian would prefer not to get embroiled in a discussion of knowledge, but to instead consider the table from the perspective of what likelihood ought rationally be assigned, given the evidence, limitations and goals of the agent.

2.4   REFLECTIVE KNOWLEDGE

Reflective knowledge is conscious knowing. Of the creatures that can know, not all have the capacity to reflect upon their knowledge. Reflective knowledge that p *requires* belief about our belief that p, and it emerges from thinking carefully, such as persevering through counter arguments, skeptical concerns,

differing contexts, and changed circumstances. Although reflective knowledge that p requires belief about our belief that p, it does not require an infinite set of justifications about one's belief about one's belief that p.[40] Reflective knowledge is also the domain of highly abstract thinking. If a person knows the truth of necessity, causality, identity, and goodness, they understand them in the domain of reflective knowledge. Mathematical, metaphysical, epistemic, or conceptual truths are established in the higher order thinking of reflective knowledge. Rationalists argue that reflective knowledge is connected with the *a priori*—beliefs arise from reason, insight or intuition, rather than sensory experience—but there is a growing literature (see Kornblith, 2012) that naturalizes reflective knowledge, building on the empiricism of Locke and Hume. One can agree that the mind has levels of capacity and sophistication without supposing that the origins of these capacities lie beyond contact with the world.

The apparatus of Sosa's reflective knowledge includes mental faculties traditionally associated with rationalism rather than reliabilism—i.e., "reflective acquisition of knowledge is…like attaining a prized objective guided by one's own intelligence, information, and deliberation" (Sosa, 2009a, p. 142). Sosa's vision for reflective knowledge is that it incorporates higher order cognition to improve epistemic virtue (i.e., knowledge) via experiences in the world. However, even reflective knowledge "tend[s] to rise only to a Moorean commensense level, below the sophistication of any Cartesian epistemology supernaturalized, or any Quinean epistemology naturalized." (p.64)

---

[40] However, I do discuss the problem of infinite regress in §2.4.1

Reflective knowledge is built up over time and experience. The more aware we are of the reliability—or unreliability—of our faculties, the greater the epistemic value of our beliefs. Consequently, justified trust in our intelligence, memory, and deliberative capacities ought to increase reflective knowledge according to the *Principle of the Criterion* (PC)(Sosa, 2009a, p. 139).

PC Knowledge is enhanced through justified trust in the reliability of its sources

To take a developmental example, a simple proposition expressed by a toddler is less justified than the same proposition uttered by an adult, even if both are true. The adult gains justification in virtue of their capacity to explain their beliefs within a wider theoretical or informed understanding of the world and their own mental apparatus. A reflective belief does more epistemic work than a child's belief "if it coheres properly with the believer's understanding of why it is true… and of how in which it is sustained as reliably truth-conducive" (Sosa, 2007).

Sosa's reflective knowledge has the AAA structure of apt belief with additional 'meta-aptness' drawn from higher order reliable processes. Thus, 'knowing full well' is having aptly apt beliefs. It is animal knowledge plus an "understanding of its place in a wider whole that includes one's belief and knowledge of it and how these come about" (Kornblith, 2009, p. 128). The 'meta-aptness' of reflective knowledge stems from higher order reliable processes that are themselves evaluable via the same performance model of regular apt belief. An apt performance becomes meta-apt when the agent reliably performs "competent risk assessment" based on her situation (both constitutional and circumstantial) (Sosa, 2011). Therefore, reflective knowledge is obtained only if

one can rule out every possibility that is incompatible with knowledge. This is the *principle of exclusion (PE)*:

> PE "If one is to know that *h*, then one must exclude (rule out) every possibility that one knows to be incompatible with one's knowing that *h*" (Sosa, 1997, p. 411)

I know that I have a hand because I can rule out any possibility that is incompatible with my hand being here, attached to my body. The reflective knower has the capacity to defend his or her belief against any rational reason to reject it:

> …in order to know full well that p, one must be able to "defend it in the arena of reflection": one must be able to view oneself as meeting every condition that one recognizes as required in order then to know that p; or, alternatively and to the same effect, one must be able to exclude justifiedly any possibility one consciously recognizes to be incompatible with one's then knowing that p"

> and

> "one does not attain high-level [reflective] knowledge, when one consciously wonders whether one does know, unless one is able to say yes…with adequate justification". (A. Some consequences of two principles, Lecture 6. The Problem of the Criterion, Sosa, 2007).

Sosa is careful to acknowledge that one does not need to be consciously thinking of all these possibilities at the moment a belief is occurrent, only that a person "would be able to defend [it], no holds barred, were it cast in the arena, perhaps by a hypothetical skeptic."[41] Knowing full well requires being disposed to defend one's beliefs by being able to explain how apt beliefs were formed with the guidance of one's second-order apt perspective. Sosa explains,

> Diana's performing full well requires not only that she host an apt perspective that constitutes her apt risk assessment. It requires also that she choose whether to act under the guidance of that perspective, and not, say, by a coin toss[42].

---

[41] Ibid.
[42] Personal communication August 15, 2013.

Additionally, rather than coming up with explanations post hoc, the knower has thought of all skeptical concerns *before* being called upon to defend themselves. Knowing full well entails having the disposition to reject rational alternatives under skeptical pressure.

According to the *principle of exclusion*, one must rule out every possibility that is incompatible with one's knowing that *h*. But what precisely is meant by 'every possibility' and how does our understanding of this affect the demands placed upon knowing full well? Surely 'every possibility' suggests that we do not have to rule out simply *conceivable* circumstances We ought not need to have dispositional readiness to defeat totally improbable, yet conceivable circumstances such as the notion that our hand is a holograph or that our memories are being systematically manipulated by a miraculous entity that does not obey the laws of physics. If 'every possibility' rules out simply conceivable worlds, then, what level of possibility must be excluded to know full well?

Any very likely scenarios ought to be ruled out. I take likely scenarios to be close possible worlds where mundane differences result from every day choices, such as wearing a different sort of outfit, choosing a different route to work, deliberating more carefully before making a decision, or coming to a different belief about an important matter through different evidence. For example, I know that I ate porridge for breakfast today, rather than eggs on toast, even though I had eggs on toast earlier in the week. Apart from these mundane worlds, more remote worlds that obey the laws of science ought also be excluded. For example, knowing full well means having a sufficient grasp of cognition to know that

sometimes dream beliefs and experiences can be present and vivid to the waking mind and can cause minor confusion before being assessed and quarantined. Thus, excluding dream beliefs is important when considering one's mental state over breakfast. In combination, these constraints (ruling out close possible worlds and more distant, science-compatible worlds) mean that the possibilities that must be ruled out are only those that *cohere with the rest of a person's beliefs about the way the world behaves and their own cognitive system*. The new principle of exclusion via coherence (PEC) is:

> PEC: if one is to know that h, then one must exclude (rule out) every possibility that one knows to be incompatible with one's knowing that h that coheres with the rest of one's beliefs"

Reflective knowledge requires being able to defeat possible counterarguments, but how much must one consider before knowledge is assured? Might there not be an infinite regress of cohering justifications that block knowledge attribution?

2.4.1 THE RISK OF INFINITE REGRESS

The Pyrrhonian dilemma refers to the problem of offering a justification for beliefs that does not lead to an infinite regress of justifications. That is, if a knower's beliefs are justified by providing reasons for their beliefs, do we not also need justification for those reasons?—and then justifications for those higher level justifications?—and so on. The Pyrrhonian dilemma for Sosa (and other coherentists) is an infinite regress of justification from animal knowledge to reflective knowledge and to even higher reflective knowledge and so forth (BonJour, 2003, pp. 197-198). Pyrrhonians argue that "any attempt to move

beyond foundations only misleads us into circles or regresses, viciously either way." (Sosa, 2009a, p. 145) I argue, however, that degrees of knowledge can acknowledge and diffuse iterative concerns. The regress of aptness and meta-aptness is not a threat to virtue epistemology, but an explanatory tool to describe different levels of competence and performance domains. The fact that aptness and meta-aptness could theoretically continue iteratively may present an *inevitable*, but not *vicious* circle.

In this context, a vicious circle implies a never-ending search for the truth without ever knowing truth. An inevitable circle acknowledges the journey, but points out the stops along the way. The ideal knower moves through life and gains experiences, developing both more reliable knowledge and better justification for his or her beliefs in ever-increasing domains. These capabilities are not plucked from the a priori, but are honed via practice, attention, use and lessons from failure. As Sosa says, "'instinctive' reactions are still subject to fine-tuning through further practice and training" (2009a, p. 142).

Each level of justification only needs to explain an organism's apt belief within a particular context of operation. The more justificatory levels one goes up, the broader the context of operation for that being. At the ceiling, one can imagine an omniscient being would have ultimate justification over all contexts. In sum, there are many levels of justification and many levels of knowledge. The fact that these can theoretically extend is not a threat to Sosa's virtue epistemology so long as each level is described and knowledge attributed appropriately.

2.4.2 THE PROBLEM OF LUCK

Another problem for reflective knowledge is the problem of luck. Pritchard (2009) points out that Sosa's definition of knowledge allows possibly lucky shots to be considered knowledge. Consider the archer confronted with not one but two gusts of wind: first, an unlucky gust that pushes the arrow off target, and second, a lucky gust that pushes the arrow back in line with hitting the bullseye. Apparently Sosa (2007) would consider such a shot apt:

> If the act is due to a competence exercised in its appropriate conditions, its success may be due to luck in various ways. It may be just an accident that the agent retains his relevant competence, for example, or that the conditions remain appropriate. Either way, the act fails to be safely successful, since it might too easily have failed, through lack of the required competence or conditions. It might still be apt, nevertheless, indeed attributably, creditably apt. (p. 81)

Pritchard's concern is safety. He argues that aptness cannot count for knowledge, because an apt shot *might* have hit the target due to luck. Consider a General competently sending out orders down the chain of command to achieve a strategic outcome. Mostly the message is received accurately, thus is apt. However, there might be circumstances where errors occur. It is easy to imagine a pair of communication mistakes that accidently result in the correct message arriving for the troops and the desired outcome achieved. When this happens, the General was partly responsible for the message's success, but his competence is exhibited in the strategic vision, not in ensuring reliable, clear communication. Pritchard's point is that because the General's capacity to get messages to the troops is unreliable, it is not apt even in the cases where it works. Aptness must come from consistent reliability. Thus, for Sosa to maintain his AAA structure against the critics, he may need to sure up his notion of reliability.

The earlier example of the double gust of wind (and indeed Sosa's own version with a single gust of wind), is useful because it is analogous to belief and mental processes, yet how analogous are these cases really? Consider the period of time between the archer releasing the arrow and it hitting its target, during which the archer has no control or influence over the trajectory of the arrow. (Note that the same 'no control' period exists for the General above). The opportunity for lucky gusts of wind to upset the outcome is when the archer has no control. While this is true for arrows, is it true of beliefs?

Beliefs are never loosened from man's quiver. Beliefs are formed and deployed as time passes and circumstances change. Information and influences bear down on beliefs over time and are reacted to accordingly. The correct analogy for the archer is more like being both the releaser of the arrow and the radio operator (able to alter the trajectory to account for the wind changes). Or the archer holds a special device that can spurt puffs of air at the arrow to recorrect for freak gusts, or designs an arrow that can self-correct after being battered by freak gusts.

The example of modern missiles also illustrates this point. Humans have designed the Tomahawk subsonic cruise missile to self-correct by comparing the terrain beneath it to satellite-generated maps stored onboard. If its trajectory is altered, motors will move the wings to counter unforeseen difficulties on its journey. The tactical Tomahawk can be reprogrammed remotely mid-flight to a different target using GPS coordinates stored locally. Similarly, a human being is both an animal agent with implicit systems (that course-correct and reevaluate

circumstances for each moment in time), and a reflective agent that aims and redirects. So, how do the Tomahawk engineers' and operators' competencies play a role in the success or failure of the missile to hit its target? Each gust of wind prompts automated reaction. If part of the guidance system fails, human decisions will affect how well the missile flies. The role of luck is minimized by the constant engagement of agents within their own mind and their environments.

Another point is that many lucky outcomes are predictable, such as gusts of wind (consider the level of obsession with wind character, strength, and direction amongst professional golfers). Golfers would not attribute the full success of a shot that was buffeted by a double gust of wind to a player, but then golfers do not attribute any success entirely to skill. Everyone implicitly acknowledges the role of luck in any success. One of the pleasures of games is that they involve both luck and skill. There is *always* some luck for archers, golfers and others striving to hit a target in intrinsically imperfect conditions. Is there always the same balance of luck for human beliefs? Even if luck often plays a role in belief formation, surely some beliefs and belief-generating processes are certain in a particular context? See Chapter 4. For further comments on this issue.

I certainly agree that humans do not *know* through luck, but the onus is on the critics to draw intuitions from appropriate analogies. Consider the Tomahawk missile: Even if wires short circuit, a human operator can intervene; conversely, if a human operator is waylaid, the missile can self-direct to the target. Similarly, virtue epistemology has layers of epistemic 'course correction.' In any case, supposing a better luck example can be imagined (and I do not doubt

philosophers' capacities to create some outlandish example involving angels or other supernatural forces), should that upset virtue epistemology?

Part of the reason why credences need to play a greater role in epistemology is that instances where knowledge does not obtain—yet competent processes are deployed—are often *no big deal*[43]. Knowledge is the ideal epistemic state, but epistemic states where luck defeats knowledge may leave behind another valuable justified epistemic state, plus a feedback opportunity to increase knowledge (or the probability of knowledge) for future situations. Alternatively, the lucky circumstance may be so unlikely that the absence of knowledge in the case specified is inconsequential for future actions. Likely luck can be understood and prepared for. If the tomahawk missile short circuits, engineers will investigate and reinforce the relevant mechanisms for future flights. Reliable processes are those that create knowledge, but also improve the odds of future knowledge in different conditions. Instances of knowledge are valuable in that they inform the agent and those around them of the scale of their competencies. However, there is no shame in justified credences or accidental true belief when treated appropriately by the agent.

2.4.3 KK PRINCIPLE

The KK principle is the idea that to know, one must know that one knows. The Cartesian characterization was that there was cognitio (knowledge) and then scientia (higher, reflective, enlightened knowledge). Scientia is a particularly

---

[43] I do not mean that all circumstances where knowledge does not obtain are inconsequential, merely that often we believe something that is almost true, such as we believe that our car is running out of gas and we only have 25 miles before our fuel tank is empty, but in actual fact, our car only has 20 miles of fuel left in the fuel tank. Presuming we prioritize getting more gas, the failure to know is not going to affect our ability to get to our destination.

comprehensive and coherent form of knowledge—knowledge of ourselves and our place in the universe—that stems from persistent reasoning. Wilfrid Sellars (Sosa, 2011) summarizes the intuition well: real knowledge only occurs when it is "…*in some sense* recognized by the person whose belief it is." Say I know *C*—the cows will be fed at 5pm. The KK intuition is that if I know C, then I also know how I came to have this knowledge, such as through reliable testimony from the farmer, or through reliable perceptual experiences. The KK intuition derives from knowledge defined as 'justified true belief,' where 'justified' refers to some sort of rational or conscious appreciation for the knower's state. The intuition is that, in virtue of being in such a reflective state, the knower is aware and thus knows when she knows. The knower can sense when knowledge obtains. Perhaps the knower has a quiet confidence, or a firm grasp on many factors relating to their current experience?

At this point I would like to introduce two siblings, Diffident and his sister Dauntless[44]. Diffident and Dauntless are both highly experienced sailors, but they have different temperaments. Diffident tends to be more cautious with regards to his beliefs, setting his credences closer to indifference, even if he has evidence that justifies a high level of belief. Dauntless tends in the opposite direction and is overly confident with her evidence. If her evidence suggests a modest belief, she tends to push her credence further towards 0 or 1. Suppose the siblings set off on a circumnavigation of the globe in a 39' yacht. They take turns at the helm; making decisions differently, and these decisions have varying outcomes.

---

[44] The character 'Diffident' stems from Sosa's Knowing Full Well (Sosa, 1993). Sosa introduces him and his counterparts 'Normal' and 'Assertive' in the following way: "How important epistemically is the distance from the actual to the ideal?" For the sake of narrative flow, I have chosen the name 'Dauntless', rather than 'Assertive'.

We can imagine Dauntless knowing the right moment to jibe based on lining up two conspicuous trees on the shore behind one another. Her knowledge extends beyond physical knowledge to pull the rudder firmly portside—she knows that she knows this peculiar fact about the landscape and that safe passage is found between the shallow reefs beneath her. Her confidence stems from memories she has of navigating this in precisely the same way in the past during different winds, different swells and different crew. She knows that she knows.

Conversely, Diffident sometimes knows when he doesn't know. Say he makes an off-the-cuff comment that the QE2[45] can fit beneath the Sydney Harbour Bridge. Dauntless scoffs and asks him to put a modest wager on his claim. The mere discussion of money makes Diffident uneasy. Maybe he was wrong? His unease turns him off the idea and he refuses the bet until he feels confident that he knows. Here the feelings of dread count as evidence against his initial confidence. Upon checking their facts, it turns out the Sydney Harbour Bridge has a clearance of 49m[46] and the QE2 is 2 metres too tall, rising 52.1m[47] above sea level. Diffident reviews his memories assiduously to find some explanation for his own confusion. He thinks he must have originally heard the story about the QE2 as a boy and forgotten the details. Worse, he reversed the pertinent fact about the ship—that it was marginally too big for the bridge. Diffident's urge to check whether he knew forced him to reflect and revise his beliefs. Virtuously, Diffident quickly assimilated truthful propositions in light of

---

[45] The Queen Elizabeth II (QEII) was a famous ocean liner built in 1968. Her predecessor, the QE, was an infamous Allied troop carrier in World War II, who, along with her sister, the QM, were said by Winston Churchill to have shortened the war by one year. http://trove.nla.gov.au/ndp/del/article/26074619
[46] http://en.wikipedia.org/wiki/Sydney_Harbour_Bridge accessed 7 March 2013.
[47] http://en.wikipedia.org/wiki/QE2 accessed 7 March 2013

skeptical pressure and disconfirming evidence. Do all people respond appropriately to skeptical pressure or disconfirming evidence? I will consider both in turn.

The problems with KK are twofold: 1) You can believe that you do not know, but still know; and 2) You can believe that you know, but you do not know. Consider Diffident and Dauntless sailing past sunset to reach safe harbor behind Cedar Bay before a storm batters the North Carolina coast. Dauntless claims that she saw two flashes, 15 seconds apart, indicating that they have passed Cape Lookout lighthouse.[48] She believes her calculations and continues to believe them strongly—even when challenged by Diffident—digging in her heels. It is critical to get their trajectory correct to get a safe night's sleep, but a strong wind and a broad reach means they have sailed quickly away from perceptible additional lighthouse flashes. As it turns out, Dauntless either misperceived or misremembered the pattern of lighthouse flashes she saw earlier. Perhaps she mistook Diffident's torch light—which she saw in the corner of her eye as she focused on trimming the mainsail—for a lighthouse flash? Perhaps she did not actually count the seconds, and instead estimated them poorly based on subjective time dilation (i.e., 7.5 seconds *felt* like 15 seconds).[49] Either way, their actual position is just north of Cape Hatteras lighthouse (which flashes every 7.5 seconds). Because of her mistake, their vessel is in open water and unable to berth safely. Dauntless does not know where they are, but, she believes that she knows and she knows that she believes that she knows. She can give good reasons for

---

[48] http://www.us-lighthouses.com/faq.php accessed 8th March 2013

[49] Subjective time dilation is documented during 'action preparedness', that is when a professional ball game player sees the ball coming towards them and must focus and plan their response (Hagura, Kanai, Orgs, & Haggard, 2012)

why she feels she knows what she knows, such as her experience sailing and navigating using lighthouse flashes in the past. The key point for our discussion is that her *feeling* of confidence is the same regardless of whether she does or does not know. The *feeling* of knowing is a fallible guide to knowledge.

The opposite circumstance can apply too. That is, someone can know something, yet not know that they know it. Diffident—being more cautious generally—knew that they were north of Cape Hatteras Lighthouse. Perhaps he knew this because he accurately counted the light flashing 7.5 seconds apart, or because he perceived swirling black and white stripes across its column, rather than the black and white diamond shapes that define Cape Lookout? Either way, he does not feel confident. He knows where they are, but he does not know that he knows this. Thus, he assents to Dauntless' confident dismissal of his skeptical concern. The *feeling* of not knowing is also a fallible guide to not knowing.

What does this mean for Sosa's distinction between animal and reflective knowledge? Animal knowledge does not require KK to be traditionally defined. Sosa's 'knowing' equates to apt belief, but no introspective connection to this aptness. This means that I can know that I have a hand without any reflection about it. Of course, if knowledge is defined as 'apt belief' as a performance, then KK means having an apt belief about apt belief. This meta-aptness is reflective knowledge—a beneficial and valuable sort of knowing, but not one necessary for knowledge.

Having knowledge requires an agent to have utilized reliable processes in an appropriate environment to achieve true beliefs. A reflective agent capable of

introspection and testimony will intuit whether they are in a state of disbelief, withholding, or belief, and they ought to have a sense of the *degree* to which they are in any of these three states. However, knowing whether they are in a *knowledge* state may be elusive to them. An agent may not know if or how their beliefs about beliefs are justified, or indeed whether they are true. This is okay, because agents do not need to have knowledge to be credible informants. Remember that for an agent to be virtuous, they just need to be reliable at gauging the likelihood of their non-belief/belief state.

A "philosophically satisfying" account of reflective knowledge ought to be one where one's reasons for trusting the faculty do not in themselves depend on the faculty. The beauty of reliabilism is that it does not rely on introspection to validate its operation. Animal knowledge is easy this way, but reflective processes also work regardless of how one understands how they work or can justify them to an enquiring skeptic. Our evidence of their success is measured by performance success, which is evaluated through our remembrances calibrated against testimony and epistemic artifacts in our external environment, such as written evidence, alerts and alarms, ordered physical spaces, and photos and videos. Our responsibility as reflective knowers is not to have a theoretical understanding of the mechanisms of reflective knowledge, but rather to adjust our trust in these capacities according to their performance context and success.

Consider Sosa's example of a hunter. A hunter does not need to know the physics of muscular-skeletal function in order to adjust her shot based on her perception of a change in wind conditions, visibility, or tension in her bow.

Similarly, the reflective knower need not know the levels of interacting mechanisms that produce human understanding. A reflective knower's responsibility is to evaluate the context of knowing and adjust her confidence and her beliefs accordingly. The reflective knower's reasons for trusting her faculties are limited to being able to report on the level of success of their prior performances.

Sosa (Ch. 3 2011) claims that there is a *sufficiency intuition* for knowledge, meaning that once a particular justificatory threshold has been reached, it does not matter exactly how confident a person is in their belief past that point. His motivation is to create conditions for knowledge without the forbidding demands of certainty. An uncertain belief, believed with sufficient evidence to manifest knowledge, is apt (i.e., not flawed) (2011). Sosa gives the following example:

> Suppose that given the evidence at his disposal, Diffident should be extremely confident, while yet his great intellectual caution makes him much less confident. His belief may then still be highly justified epistemically, with the sort of full justification relevant to whether one knows. Diffident's belief could then be justified, surely, even if he could properly be much more confident than he is, with justification to spare. Compare Normal, who has much slighter evidence than Diffident on the question at issue. Sufficiently weightier evidence could make Diffident better justified, and might even trump the fact that Normal's actual degree of confidence is perfectly aligned with his ideal degree. (p.36).

Sosa's point is that as long as Diffident's reasoning pushes him correctly past the threshold to belief (or disbelief), perhaps it does not matter precisely what degree of confidence he has in his beliefs? The same point can be made for the overconfident Dauntless.

Is it epistemically *insignificant* for an agent's confidence to be distant from the ideal? I argue that while a single instance of variance may not have

detrimental consequences for a particular performance, it is epistemically very significant in the longer term. Subtle degrees of confidence can have long-term impacts on epistemic success—thresholds are not the only things that matter. The Bayesian defense of formalized degrees of confidence or belief is that beliefs ranging across many degrees of confidence are rational components of decision-making. Failure to discriminate gradations of belief carefully will lead to a guaranteed loss for an agent over the long term—incurring a *Dutch book* (Parmigiani & Inoue, 2009).

2.4.4 ANIMAL KNOWLEDGE TWICE OVER

A number of commentators (Goldman, 2008; Kornblith, 2009) have asked how reflective knowledge adds anything epistemically to what has been achieved via regular apt belief. Goldman asks, "if a first-level of reliability and non-accidentality is inadequate to achieve genuinely human knowledge, why does an added layer of the same deficient stuff turn low-grade knowledge into high-grade knowledge?" (Goldman, 2008). Goldman cannot understand how the iterative mechanisms that produce animal knowledge could produce differentiated justification for higher order beliefs. Hilary Kornblith claims that Sosa's reflective knowledge cannot gain justification beyond that established by animal knowledge unless Sosa adopts traditional rationalist (i.e., non reliabilist) justificatory apparatus. In the end, he considers reflective knowledge to be just 'animal knowledge twice over.'

Kornblith uses the case of 'Norman' to explain his concerns about using the same mechanisms of animal knowledge for reflective knowledge (2009). Norman

has a reliable capacity to form beliefs due to clairvoyance. He is able to form true beliefs about the future—for example, the US President's whereabouts at a particular time psychically—but he is unable to defend this capacity. Additionally, he "does not have evidence of having a reliable clairvoyant capacity, and he has no beliefs at all about whether he has such a capacity" (BonJour, 1985). This supposedly qualifies as having animal knowledge, without reflective knowledge. Kornblith goes on to imagine 'Aptly apt Norman.' Aptly apt Norman has apt beliefs about his apt beliefs. Without being able to offer an explanation, he simply expresses his belief—"I just believe my belief about the President's whereabouts is apt," (Kornblith, 2009)—when asked about it. Kornblith argues that this explanation does not defend either the first order or second order belief. Thus, he states that "we cannot legitimately make the move from 'aptly believed to be apt' to 'defensible'" (p. 129).

I question whether Kornblith's hypothetical scenario is a suitable test of Sosa's views. Is Aptly apt Norman really 'aptly apt' if he has no evidence or belief of having a reliable capacity? Animal knowledge is forged from feedback within one's environment. Second-order beliefs about the aptness of first order beliefs are also formed using precisely the sort of evidence that Norman lacks. Without feedback about our second-order beliefs, we are not in a position to defend them. Let me be clear, I do not mean that Norman ought to have feedback about the structure of his cognitive capacities, but rather that he does not have any feedback about the aptness of his performances. Norman fails to have aptly apt beliefs because he cannot report on previous performances, not because he is

ignorant of the mechanisms that create his beliefs. Norman's belief, though true, is not integrated with any other beliefs. Thus, it is not surprising that Kornblith thinks that Aptly apt Norman's self-examination confers no defense—it doesn't.

Yet, the failure of Aptly Apt Norman to have reflective knowledge does not bring down Sosa's theory because it is unclear how Norman would believe that his beliefs about the president's whereabouts are apt without any further explanation. If he is a normal sort of person, then beliefs of the sort he is experiencing would lack the right type of justification to count as aptly apt—or even apt. Now, if he had some sort of theory about clairvoyance (even its possibility) and this was a probable belief, then he might have some reason to suppose that his believing the President's whereabouts might be apt. However, there is no indication that he does and, in any case, there is no clairvoyance. The difference between BonJour's example and real cognition is that an organism gets continual feedback from the impact of their animal beliefs via the environment or social interaction. Animals build confidence via learning and experience. BonJour's clairvoyant has no evidence for his capacity[50].

There are ways that animal knowledge and reflective knowledge can be built with iterations of similar processes to produce differentiated results. Reflective and animal knowledge are both forged in a Bayesian fire. Just as a blacksmith can create a wrought iron poker or decorative gate, the *instantiation* of animal and reflective knowledge is quite different. Animal knowledge

---

[50] One can imagine Kornblith responding that even if humans do not have Norman-style clairvoyance, it is logically possible that some beings do and that these beings have clairvoyant access to their first order competences sufficient for reflective knowledge (Ernest Sosa, personal communication, August 12, 2013). I agree that it is logically possible that such a capacity exists, but evolved creatures of earth are not so endowed, thus the relevant study of knowledge must be catered to human (or higher order animal) knowledge-generating competencies.

mechanisms are largely low-level, bottom up processes that systematically sort and categorize sensory experiences into basic beliefs and adaptive behaviors. Reflective knowledge mechanisms are largely high-level, top-down processes that systematically analyze and reconsider beliefs drawn from animal processes. Reflective knowledge yields 'enlightened discovery' (Sosa, 2009a, p. 146), which makes an agent better at evaluating beliefs and honing expectations and intuitions. Using and developing reflective knowledge improves the efficiency with which an agent achieves its goals. There need not be anything magical or 'a priori' about reflective knowledge—we build reflective capacity just as we develop other reliable processes. Cognitive scientists (Gallistel & King, 2009) and decision-theorists (Bovens & Hartmann, 2004; Gwin, 2011; Hajek & Hartmann, 2009; Pearl, 1985; Talbott, 2011) suggest that Bayesian processes *must* underlie knowledge acquisition at all levels.

Still, even if BonJour's Norman is not a good example, it is reasonable to wonder how reliable animal knowledge could become rationally defensible reflective knowledge through the application of 'the same stuff.' How could 'dumb' Bayesian processes produce enlightened Cartesian 'scientia'? Is this just connectionism (Elman et al., 1997) or dynamic systems theory (Thelen & Smith, 1996) redux? A new propositional approach using hierarchical, Bayesian processes (Gopnik & Wellman, 2012) may demonstrate how "meta-aptness imports some knowledge and understanding... of how one's first-order belief attains success" (Sosa, 2009b). A successful account of reflective knowledge must explain how a child develops adequate abstract, causal, or counterfactual

knowledge from iterations of animal knowledge. Thus, as I argue, the same Bayesian framework that makes sense of animal knowledge can be layered to function in particularly complex ways.

2.5   THE THEORY THEORY OF REFLECTIVE KNOWLEDGE

Sosa's account of reflective knowledge coheres with work in cognitive psychology on how people develop theories of how the mind works—the so-called 'theory theory.' The theory theory attempts to explain higher order belief revision by proposing that all humans, from babies to adults, come to understand the world abstractly by forming hypotheses and then rigorously testing them and updating them in light of evidence. The 'theory theory' supposes that children and adults develop or construct hypotheses of the world and alter them according to new evidence.

Theories have a distinctive structure involving coherent, abstract, causal representations of the world, including unobservable theoretical entities. Higher-level theories describe kinds of entities and relations more generally, rather than particulars within a domain. Theories have particular cognitive functions including predicting future events and interpretations of evidence, and they allow counterfactual inferences. Counterfactual inference might relate to what might have happened in the past, or to what would happen if something intervened in the present to affect the future. Theories that have been studied empirically include intuitive theories of one's own mind and others (Gopnik & Wellman, 2012), the biological world (Gopnik, Meltzoff, & Bryant, 1997), the physical world (Gelman, 2003) and social world (Baillargeon, 2008). These theories

enable individuals to understand and predict phenomena in the world via counterfactual inference.

As promising as the 'theory theory' research program has been, it has been quite vague on how theories are represented or what learning mechanisms underpin them. Rational constructivism (Xu, Dewar, & Perfors, 2009) uses the framework of probabilistic models, specifically Bayesian learning, to defend a new version of the 'theory theory' of cognitive development. The marriage of the 'theory theory' with probabilistic models seems a promising way to map the development of reflective knowledge. New work being done on hierarchically nested probabilistic models (HNPM) shows how complex thoughts can be achieved through iterations of the same justificatory processes that underlie basic probabilistic processes. HNPM are structured, describing the relations between models and patterns of evidence in rigorous ways.

HNPM show that higher order theories (e.g., about abstract ideas) can become inductive constraints on the interpretation of lower level theories—what Goodman (1983) describes as an *overhypothesis*. Higher order theories remain dynamic because they are updated in light of new evidence, unlike modules or 'core knowledge' (Spelke & Kinzler, 2007) that remain fairly static. Unlike associationist structures, dynamic theories change at a high, broad level, not just at the level of particulars. It turns out that children and adults continually conduct informal experiments (exploratory play, observation, imitation, and intervention) in order to evaluate and revise their theories of the world. Cognitive development depends on statistical information about the probabilistic contingencies between

events. Models help individuals predict and hypothesize the abstract structure of the world from perceptual or low-level inputs.

The capacity to form abstract theories from low-level data is known as 'the inverse problem.' In vision, the 'inverse problem' means inferring properties of three-dimensional objects from flat retinal images. The 'inverse problem' for theory change means inferring the causal structure of the world from observing events. The traditional problem for these sorts of inferences is that the set of possible hypotheses commensurate with available evidence is almost unlimited. Indeed, it is the intractable nature of the problem that has led theorists to suggest that theories of reality must be innate (Pinker, 1984).

The difference between traditional modeling and newer, probabilistic modeling is the impact of Bayesian inference on reducing the scope of plausible hypotheses. Bayes' rule limits the range of hypotheses under consideration by making a prediction about the probability of each hypothesis, such that one need only consider seriously the most likely under a Bayesian interpretation. That is, Bayesian methods allow individuals to determine the probability of the possibilities to reduce the range of plausible hypotheses. Gopnik & Wellman (2012) give the following example:

> Suppose Mary is travelling, and she wakes up with a terrible pain in her neck. She considers three possible hypotheses about what caused the pain: perhaps she has a clogged carotid artery, perhaps she slept in an awkward position on that wretched lumpy mattress, or perhaps it was that dubious lobster she ate last night. She goes to WebMD and discovers that both a clogged artery and awkward sleeping position are much more likely to lead to neck aches than bad shellfish—neck aches have a higher likelihood of occurring given a clogged carotid and awkward sleeping position than they do given ingestion of bad shellfish. In fact, Mary reads that clogged carotids always lead to neck aches—the likelihood of a neck ache given a clogged carotid is particularly high. Should she panic? Not yet. After all, it is much less likely to begin

with that she has a clogged carotid artery than that she slept awkwardly or ate bad lobster—awkward sleeping positions and bad lobsters have a higher prior probability than severely blocked carotids. If you combined these two factors, the likelihood and the prior, you would conclude that a bad night on the lumpy mattress is the most likely hypothesis.

Eventually though, enough evidence could lead you to accept even an initially very unlikely idea. Sufficient additional evidence (the ache persists, an X-ray shows blockage) might indeed lead to the initially unlikely and grim diagnosis of a clogged carotid artery. This gives Bayesian reasoning a characteristic combination of stability and flexibility. You will not abandon a very likely hypothesis right away, but only if enough counter-evidence accumulates (p.1088).

Still, how effectively do Bayesian solutions reduce the set of possible hypotheses? By itself, Bayes' rule is not sufficient. What is needed is a more complex application of Bayesianism, also known as 'Bayes nets.' Bayes nets are coherent, systematic, hierarchical sets of hypotheses, nested and interacting with evidence on different levels to interpret complex events using finite cognitive resources. These complex applications lead to HNPM.

HNPM can account for multiple levels of knowledge, including: 1) abstract generalizations relating to higher level principles, 2) specific theories about a set of instances, and 3) particular experiences. HNPM explains how abstract generalizations arise from specific theories that are, in turn, learnt from particular experiences. What is strikingly different about these models (compared with classic empiricist, foundational accounts of knowledge acquisition) is that abstract generalizations can precede specific ones, as I explain below.

Suppose many bags of marbles were placed in front of you and your job was to identify the color of the marbles in each bag (Gopnik & Wellman, 2012). An experimenter takes a red marble out of the first bag and asks you what color you think the next marble will be. She then goes on removing marbles from that bag, pausing between each one to ask the same question, and each time the marble

is red. The experimenter then repeats the procedure with the second bag of marbles. This time a succession of blue marbles appears. You will quickly assume that all the bags of marbles contain only one color. Predicting that the contents of any bag of marbles will match the color of the first one is an abstract generalization—an *overhypothesis*. Learning this *overhypothesis* precedes learning about the contents of bag 3 (for eample, that all marbles are purple). In fact, hierarchical Bayesian models make it just as easy to learn causal structures at several levels at once. As Gopnik & Wellman (2012) explain, "Probabilistic hierarchical Bayesian learners thus learn abstract structures alongside and even before the specifics that those regularities subsume". These abstract structures can then be used to limit the scope of hypotheses considered at a range of levels. Reflective knowledge consists of beliefs that cohere across multiple levels of explanation.

Work on HNPM is still in a very preliminary stage, and it is not yet clear how radically new hypotheses are generated from data. One possible addition is 'Quinean bootstrapping,' where linguistic structures and analogical reasoning affect conceptual insight (Carey, 2009).[51] Another suggestion is that new causal hypotheses are formed in exploratory, pretend, and imaginary childhood play (Buchsbaum, Bridgers, & Weisberg, 2012). Indeed, traditional adult methods of counterfactual reasoning and conjecture most likely play a role in the

---

[51] Although this suggestion (like all others) cannot explain which analogies—of the infinitely available numbers—are picked as suitable. Note that this is another example of the relevance problem that plagues all accounts. As Fodor (1983) explains, "how … does the machine's program determine which beliefs the robot ought to re-evaluate given that it has embarked upon some or other course of action?" (p.114).

development of reflective knowledge. However, cognitive science is yet to untangle the computational, algorithmic, or instantiation of these capabilities.

In sum, reflective knowledge is not, as some have argued, merely iterations of defective animal knowledge. New work in HNPM suggests that many aspects of higher-level knowledge can be created iteratively from low-level processes. Newer probabilistic models can restrict the scope of hypotheses considered and explain how multiple levels of knowledge can be learnt at once. Brute, reliabilist processes can generate reflective knowledge if they yield cohering beliefs across a broad explanatory domain.

How then should these findings affect our definition of reflective knowledge? I argue that a further condition ought to be placed on PEC that allows multiple hypotheses to be considered simultaneously and to be considered possible within the epistemic state of the knower. The principle of exclusion via coherence of multiple hypotheses PECMH is:

> PECMH: if one is to know that h, then h must be the most likely amongst a set of plausible hypotheses h_m to h_n that cohere with the rest of one's beliefs at lower likelihoods.

Reflective knowledge requires the knower to believe the hypothesis with the greatest evidence and the greatest coherence with other beliefs, and to hold other, less likely (but possible) hypotheses within consideration. A Bayesian virtue epistemology values both knowledge and credences. Epistemology becomes not just the study of justified true belief then, but also the study of the processes of belief revision in response to confirming or disconfirming evidence.

Justification arises from the apt performance of reliable processes and their coherence with other beliefs.

A significant problem still exists for a Bayesian virtue epistemology, however. How do we know that reliable Bayesian processes actually connect with the true structure of the world, rather than just an instrumentally valid facsimile? This skeptical concern is known as the 'new evil demon' problem, which I address it in the next section.

2.6   NEW EVIL DEMON

The new evil demon scenario is the idea that even though we feel as though our cognitive processes (perception, memory, and inference) are reliable, they are not because the world is not as it appears. Instead, the world is a fanciful hoax (Cohen, 1984). The main thrust of the argument is that there is no method a reliabilist could rely on to break out of the hoax, because the reliabilist toolkit consists entirely of mechanical processes that draw on (false) information from the external world to justify his beliefs. The worry is that a successful capacity to operate in the world does not necessarily mean that our beliefs correspond to an underlying reality (Comesaña, 2002). I offer two natural worlds (i.e., probable, consistent with natural laws and existing knowledge) to examine the new evil demon problem:

1. A high epistemic standards world (HES), where accurate recollection of past events is socially very highly regarded:

2. A low epistemic standards world (LES), where recollection is either not valued, or is valued as an exercise in reinforcing a collective identity, but with less concern for truth.

Anthropologists have found that autobiographical memory (episodic and semantic memory about oneself) develops quite differently depending on the cultural value placed on recollection (Fivush & Haden, 2003). Parents have a huge impact on their child's aptitude for accurately recalling past events, which they build by discussing events as they happen and then forming and reinstating the past.

2.6.1 HIGH EPISTEMIC STANDARDS WORLD

Maori children show earlier, richer, and more robust childhood memories than many cultures because of the value that is placed on the oral transmission of accurate facts about the past (Hayne & MacDonald, 2003). They live in a very high epistemic standards world (HES). In HES, Maori mothers create strong, accurate memories by talking with their children about events with elaboration, repetition, and evaluation. Maori mothers prompt their children for descriptive details about the larger context of the event, particular objects, and the people who were present. They include repetition, both within conversations and across conversations, to strengthen and preserve memories. They also repeatedly question and correct children to improve their accuracy. HES develops adroit children capable of strikingly apt beliefs about the past.

2.6.2 Low epistemic standards world

In low epistemic standards worlds (LES), people get very little feedback or recrimination for holding false beliefs about their memories. There might be several reasons for this. Perhaps their world is one where individualism (hence autobiographical memory itself) is not valued. Or the world is one where an individual's narrative is valued so highly that slight confabulations or even wildly implausible 'memories' are not held up to scrutiny. Leichtman, Wang, and Pillemer (2003) found that when participants from India, a highly interdependent culture, were asked to recall specific childhood memories they could only produce general memories, even when pressed. This is in contrast to participants from the United States, an independent culture, who almost all produced a memory. It is possible that within cultures that value memories, a strong personal narrative can have such high value that characteristics of stories (i.e., their dramatic effect) may be more valuable than accuracy. In LES, it is more valuable to have an engaging false narrative than to withhold due to concerns for accuracy. In LES, memories are praised for the degree to which they cohere with the collective memories being woven by the group, rather than the extent to which the events described are truthfully retold. One can imagine a child at summer camp sharing stories around the campfire. If one child describes a terrifying incident with a raccoon that gets attention, other individuals may be tempted to confabulate or construct a memory to suit the occasion rather than acknowledge a less interesting episode. In LES, intentionally and deliberately forbearing from recounting a memory due to concerns about its accuracy is a less valuable social behavior than telling a story for entertainment with no other ramifications.

LES parental guidance with regards to the value of truth might be limited or absent. In its place, invention, creativity, adaptation, and deceit are central aids to survival. A rational child growing up in these circumstances does not experience the feedback and subsequent pressure to hone their memory judgments to be reliably descriptive of the past. Instead, they speak of their memories freely and confidently, weaving them into a long temporal autobiographical tapestry that warps with each retelling. Perhaps some individuals in these circumstances can still differentiate true and false memories, but, potentially deceitful methods and stories become so dominant that the individual no longer has normal mechanisms for differentiating content. They might believe that they are reliably indicating true from false memory, but in actual fact they are deceived.

I take this LES world to be not unlike the new evil demon world—at least as it pertains to memory. Each individual in an LES world has in situ reliabilism, as described by Lyons (2012). Although it feels as though their memory processes are reliable, they are consistently deceived. LES offers a strong example against which a successful reliabilist theory must respond, and it is strong for two particular reasons. Firstly, the actual world may be a LES world or a near LES world. Normal people, or even highly trained epistemologists, might systematically endorse dramatic narrative and social coherence over truth in most instances, and this endorsement might be explicit or implicit. It might exist as a cognitive bias that is just as insidious as implicit racism in individuals who explicitly disavow racial stereotyping. Secondly, LES is a strong skeptical case because it creates a probable world for the reliabilist to solve.

Sosa (2009a) discusses a new evil demon case by the example of a young child brainwashed by a superstitious upbringing (p.36). The causal origins of the child's beliefs and inference patterns means that they are subjectively rational agents, even though they are not objectively rational. Subjective rationality is "the status that a belief has when it would survive deep reflection by the subject in the light of her deepest epistemic standards (roughly)" (p.37). That is, it does not matter how well the superstitious child reasons—or correctly takes her experiences into account or remembers those experiences—because she does not have an apt connection with reality. In the end, for Sosa, knowledge is not *just* about truth and coherence. A sane individual must also be "adequately related, causally or counterfactually [to] the objects of one's knowledge" (Sosa, 1997). The lack of an appropriate causal connection to reality means that the superstitious child does not know and is not a reliable agent.

If this is true, then has the skeptic won? How do any of us know if our beliefs and experiences have led us to true, coherent, *causally appropriate* beliefs? We need to keep in mind that animal knowledge does not require that we know (with priority) that we are properly connected, only that the correct causal connection occurs (Ernest Sosa, private communication, August 12). Similarly, if we are in a position to have reflective knowledge, the reflective processes may be opaque to us, but this is not problematic so long as the right causal connection occurs between the reflective processes and the reflective belief produced by them (Ernest Sosa, private communication, August 12). Supposing these are the conditions for knowledge. is it not still possible that we have all been brought up

in an epistemically impoverished upbringing, which gives us the collective delusion of justification?

2.6.3 THE SUCCESS OF SCIENCE ARGUMENT

Can an external justification of reflective knowledge provide sufficient grounds to defeat the new evil demon? To address this, I use the success of science argument. The success argument—also known as the 'no miracles' (Carrier, 1991) or 'ultimate' (Musgrave, 1988) argument—claims that we should believe that scientific theories actually correspond to the world because these theories have such tremendous success. Or conversely, it would be a miracle if science had the predictive success it did, if its theories did not correspond with underlying reality. The success of science is so great—so the argument goes—that achievements such as the development of antibiotics, walking on the moon and so forth would be miraculous if science had not glummed onto some approximately true theory about the theoretical entities and natural laws underpinning these successes. Similarly, agents are successful in the world because they hold true beliefs. Human survival would be a miracle if humans held beliefs about how to hunt for food, build shelter, and care for infants (let alone the quirks of our mental states) if those beliefs did not correspond to an underlying reality[52]. Science is successful because its theories are scrutinized constantly and objectively in varying conditions. Similarly, the great success of individual cognitive agents navigating their life is due to those individuals hypothesis testing their beliefs and cognitive capacities against social and environmental feedback. We become more

---

[52] I recognised that the radical skeptic will not be mollified by this argument (Ernest Sosa, August 12, 2013). The target of my concern is a conservative skeptic who raises the serious problems in the normal world such as false memories, visual illusions and so forth, rather than brains in vats and so forth.

and more able to successfully 'perform' our beliefs because we build up better and more approximately true representations of the world and our capacities within it. Reflective knowledge is justified because it has successive, successful performances, and because these performances are evaluated via external means, such as photographs, testimony, or written records. More importantly, reflective knowledge allows us to predict future performances, and to build expectations to be tested.

Sosa's reflective knowledge provides the knower with contexts of likely success just like a scientific theory provides type predictions for token instances. True scientific theories have descriptive and predictive success, particularly with regards to novel predictions. Similarly, 'knowing full well' enables an agent to make predictions about the sorts of contexts and circumstances that would put knowing at risk, and about how to mitigate against risk. Scientific realism states that only scientific theories that are 'mature' and 'non ad-hoc' count as real. Similarly, Sosa claims that individuals become better knowers over time. A human being begins life with naïve theories but develops more reliable, mature theories until an accident, dementia, or senility interrupts the proper processing of data. A reflective knower has a complex set of understandings of themselves in various environments and she evaluates her beliefs against a set of beliefs to ensure their coherence, even if the mechanisms of these reflections are invisible.

If the success of science argument is analogous to the development of reflective knowledge, then some of same problems that beleaguer that argument need also to be addressed by Sosa. Chief amongst these is the claim that success

only secures the instrumental value of a scientific theory, rather than pinpointing truth. The evidence for instrumentalism is shown by the history of theory change. Each mature scientific theory gets close enough to aspects of reality to yield limited predictive success, but sure enough, a new theory replaces the old and the cycle continues—the *pessimistic meta-induction* (Laudan, 1981). Instrumentalist concerns are particularly analogous to reflective knowledge in new evil demon cases.

It can be argued that taking the success of a theory to indicate truth is meaningless unless you know the base rate of approximately true, successful theories (Magnus & Callender, 2004; Stanford, 2010). Perhaps every theory has empirically equivalent rivals (Stanford, 2010)? However, Worrall (2007) argues against this concern by supposing that even a disproven scientific theory gets some facts right about the world and approximates the truth in an important and long-lasting fashion. He offers, as an example, the fact that Fresnel's misidentification of the nature of light still conveyed truthful facts about the structure of optical phenomena such as the relative intensities of light in refracted and reflected polarized beams. In this case, the structural facts, (i.e., mathematical equations), survived theory change (Worrall, 2007, p. 51).

Worrall (2007) argues that not all scientific progress is instrumental or interchangeable. Successful science does access at least some true facts about the world, even if the theories explaining those facts are ultimately replaced. As long as some truth or some reality is grasped by scientific theory progression, and more and more reality is uncovered with progress, instrumentalist concerns are not

destructive to scientific realism and particularly the success of science argument. Applied to reflective knowledge, Worrall's argument would mean that our theories of how our mind works might be fallible, but they're not groundless, nor are they merely coherent propositions. Our reflective knowledge actually approximates the truth and becomes more and more truth-baring and reliable as a result of experiences measured by performance success. These successes are due to the constancy of structurally true knowledge, even in the face of higher-order belief revision.

Indeed, research shows that children hold multiple hypotheses simultaneously, rather than accepting one and rejecting another. Each hypothesis is given a probability by the child and revised in light of different evidence. For example, a child's initial theory of natural kinds is morphological, but it becomes essentialist after being tested in the world. The slow, gradual change in belief "results in a characteristic series of related conceptions that forms a bridge from one broad theory to the next" (Gopnik & Wellman, 2012). The multiple-hypothesis concept is contrary to typical accounts of normal scientific progress found in Popper or Kuhn (1970), and to knowledge structures in epistemology.

Typically, scientific research is conducted within a single theoretical framework agreed to by collective majority of scientists. Typical epistemic reasoning supposes that only a single hypothesis or theory can be held at any one time about a particular proposition. However, according to the multiple hypotheses account, beliefs held probabilistically cohere in a fuzzy or gappy fashion (Alxatib & Pelletier, 2011). Such an account abides by the principle of

exclusion via coherence of multiple hypotheses (PECMH) which states that reflective knowledge requires the knower to actively consider many possible hypotheses, but to bring to prominence, and act upon, the hypothesis with the *greatest* evidence and coherence with other beliefs.

So, how do we defeat the new evil demon skeptic without begging the question? My answer is somewhat paradoxical. I argue that to defeat the skeptic we have to give in to him a little—i.e., we ought to join the skeptic in believing each skeptical hypothesis to an appropriate degree. There are many highly probable skeptical scenarios that we all face each day with regards to our memories and our cognitive faculties, and each hypothesis ought to be entertained to the degree that we have evidence for its likelihood. The answer to the new evil demon may be to relax about needing to 'defeat' the skeptic, if 'defeat' means being able to state categorically that hypothesis x or y is untrue or disproven.[53] The skeptics' laments can be heard, yet our beliefs remain justified. What seems right about the evil demon threat is that knowledge is only possible with the right causal connection to the world, with a variety of other factors, including reliable faculties using Bayesian updating. These are the methods by which we adjudicate and ultimately come to know the world.

2.6.4 Norms of Bayesian reflective knowledge

Reflective agents are active hypothesis-testers who respond to well-founded skeptical concerns and opposing evidence. They assess new information relative to existing beliefs and they evaluate the role of incoherent information, but, they

---

[53] The naturalist has good reply to the unconvinced skeptic, see Devitt (2002)

are not unduly skeptical. They trust the reliability of their faculties in circumstances that justify it (such as driving a familiar car during clement weather on familiar roads), and they can also hold multiple-plausible hypotheses simultaneously. This is an important point. Entertaining *multiple* hypotheses means that an agent is in a state between belief and disbelief with a number of theories—playing both the skeptic and the advocate at once—and it is not the same as trusting reliable faculties.

Some might be wary of this promiscuous consideration of hypotheses, supposing it to be indiscriminate, but this mischaracterizes the view. Surely it is better to speak of open-mindedness and responsible cautiousness (which Sosa (2011)refers to as withholding)? The withholding argument highlights the virtue in hedging one's bets in the absence of appropriate evidence, so long as doing so does not interfere with one's capacity to flourish in other areas. Withholding is a virtue as long as doing so does not interfere with flourishing regarding another matter. Take Sosa's (2011, p. 4) hunter example: The hunter may withhold from shooting to ensure a good shot, but, they must not withhold forever. The hunter has an obligation to find food and must, at some stage, take appropriate risk to further this aim. Similarly, a wholly rational agent ought to adopt a belief about one hypothesis or another to some degree in order to function, but they should always be ready to reevaluate as circumstances change.

Daniel Kahneman (2011, p. 154) notes that there are two essential competencies associated with explicit (i.e., reflective) Bayesian reasoning:

1. Anchor your judgment of the probability of an outcome on a plausible base rate.

2. Question the diagnosticity of your evidence.

Kahneman warns that base rates are notoriously forgotten when we consider the case in front of us and that we tend to jump to conclusions using only the evidence to hand (which he describes as the 'what you see is all there is' tendency). This tendency is absolutely prevalent. It drives an agent's intuitions and impressions, regardless of the quality and quantity of the evidence, and any associations brought to mind by this evidence will be forged into a coherent story too enticing not to believe.

In order to have a reasonable chance of strongly confirming a true hypothesis, the following additional virtues (Gwin, 2011) also need to be cultivated:

1. *Open-mindedness*: Seek alternate hypotheses because the ones we are currently considering may not be right.

2. *Adaptive*: Consider new plausibility considerations based on reevaluations, revisions, and alterations of background knowledge. Avoid fixing the prior probabilities of any hypothesis.

3. *Objective*: Create testing environments where the two hypotheses differ with regards to the objective likelihoods of possible outcomes. Keep the environments as objective and unbiased as possible with regards to one hypothesis versus another.

Gwin (2011) argues that these virtues ought to guide groups of scientists as they plan, write, conduct, and write up their experiments, but can these principles be applied to individual epistemic endeavor?

The first virtue—open-mindedness—suggests that we ought to keep considering new hypotheses. The obvious danger with such a dictate is that it would demand that we test every possible hypothesis in order to use Bayesian logic effectively. Humans must reason efficiently to meet other social and individual goals, thus it is not rational for us to consider all hypotheses. Luckily, it has been shown that only *plausible* hypotheses need to be considered (that is, hypotheses where the prior probability has been set extremely low can be left to the side of one's calculations). After continual testing, all the remaining, plausible hypotheses—including the true hypothesis H_i—will become overwhelmingly more strongly confirmed than all the other hypotheses. As mentioned in §1.7.3, ascertaining 'plausibility' is an area that pure Bayesian analysis (or Bayesian psychology) has difficulties answering, but this is a problem for many epistemic theories.

2.7 CONCLUSION

In this chapter, I proposed that apt belief is sufficient for knowledge when it is adroitly produced in the service of meeting an agent's goals, and that restricting knowledge to the domain of an agent's interest reduces the risk of infinite regress or concerns of insufficiency. I also introduced the concept of 'quasi-apt' beliefs to describe the reliable processes that create defensible and valuable degrees of belief that do not necessarily equate to knowledge. According to this view, credences that pass a certain threshold for action are knowledge if they are: a) true and b) produced aptly (true in virtue of adroit processes). The story so far looks promising. However, there are some problems that need to be addressed. Precisely how do quasi-apt and apt-beliefs interact? That is, how do Bayesian credences merge with Virtue apt beliefs? For example what does it mean if an agent is reasoning correctly using Bayesian methods and begins testing a hypothesis that has a fairly low prior, but nevertheless is true? Say, a police officer is at the beginning of her investigation of a cafe robbery and considers the hypothesis that the thief is a retiree. Now, the officer has never arrested a retiree for any crime, let alone seen a retiree convicted. But, being a diligent law enforcement agent, she considers the possibility that the retiree stole the money because he was a customer at the crime scene just prior to the incident. Still, the officer puts more weight the hypothesis that that the cash was misplaced or stolen by a barista or waitress. Now, suppose that in fact, the retiree was the guilty party and did steal the cash. What is the doxastic state of the officer when their credence for H < .5, yet H is true? Their hunch is justified because it is reliably formed using approved

and reliable Bayesian methods and it is true—the retiree did steal the money. Her hunch is not an accident or lucky. Her hunch is true in virtue of apt processes. Yet, there is something not-quite-right to say that their early intuition or hunch constitutes knowledge of H. It seems like a hypothesis with a low prior is not effective enough to count as knowledge. The intuition then is that a hunch must go through rigorous evaluation with respect to evidence in order for the agent to come to believe it to be true (i.e. to assign a high enough probability that it hits a functional threshold for belief or knowledge). Justifying the police officer's degrees of belief cannot easily be solved by either virtue epistemology or Bayesian epistemology alone. Virtue epistemology needs apparatus to manage and value subtle degrees of belief. Bayesian epistemology cannot explain knowledge attribution or its absence.

Meta-apt beliefs are higher order beliefs that constitute reflective knowledge. Reflective knowledge works iteratively through processes that are similar to those employed for animal knowledge—imparting greater wisdom and reliability through widening its context of successful operation. Joining together Bayesian processes with virtue epistemic theory means that similar justificatory apparatus can be used for animal and reflective knowledge, and that reflective beliefs may be forged from complex, hierarchical, nested Bayesian processes.

I conclude this chapter by arguing that reliable processes succeed gradually, by drawing an analogy between individual knowledge development and the growth of scientific knowledge. Under this conception of reflective knowledge, only those skeptical scenarios with high prior probabilities ought to be evaluated

and considered by a rational agent, and reflective knowers ought to exist in a constant state of reevaluation with regards to their evidence, beliefs, and hypotheses. Reflective knowledge depends on the accurate assignment of prior probabilities and the justification-bolstering impact of explanatory beliefs, both of which depend on an adequate account of coherence as a source of justification—the topic of the next chapter.

3   How coherence and coordination justify belief

"Now that you accept A and B and C and D, of course you accept Z."

"Do I?" said the Tortoise innocently. "Let's make that quite clear. I accept A and B and C and D. Suppose I still refused to accept Z?"

"Then Logic would take you by the throat, and force you to do it!" Achilles triumphantly replied. "Logic would tell you 'You ca'n't help yourself. Now that you've accepted A and B and C and D, you must accept Z!' So you've no choice, you see" (Carroll, 1895).

3.1   Introduction

Coherentism is the view that beliefs are justified in so far as they cohere with one another, and it is typically seen as distinct from foundationalism—the view that beliefs are justified from foundational sources of knowledge (e.g., perception, memory). Coherence is a relation of mutual support, consistency, or agreement between a set of beliefs (Plantinga, 1993; Sosa, 1980) that informs how inconsistent data from cognitive faculties, instruments, or testimony ought be treated (Talbott, 2011). Coherence among true beliefs offers *prima facie* evidence that a belief is true. For example, if a witness reports seeing the suspect at midnight, and this evidence aligns with the time and place of the crime, then surely the report gains some justification, regardless of the reliability of the witness? A measure of coherence can justify accepting information or explaining why information is rejected, as well as shedding light on inference to the best explanation (Weiner, 2012) and hypothesis selection (Quine & Ullian, 1970). Coherence is also important for traditional internalist virtues (e.g., logical consistency) and as such it may help Bayesianism justify the attribution of prior probabilities. However, coherentism is plagued with difficulties defining and formalizing its benefits. This chapter moves through some of these challenges and

argues for a new theory of coherentist justification that fits a hybrid Bayesian virtue epistemology.

Coherence, by itself, is an insufficient justificatory source as is suggested by the fact that a set of falsehoods can cohere without yielding the truth. A logical next step is to bolster coherence with reliability to ensure that beliefs are adequately grounded in foundational sources such as perception and memory. Weak foundationalism holds that beliefs are grounded by reliable basic processes, and that they gain justification to the degree that they cohere with beliefs from other reliable processes. Weak foundationalism is supported by both traditional epistemologists (BonJour, 1985) and Bayesian epistemologists (Lewis, 1946; Olsson, 2012; Olsson & Shogenji, 2004). Each belief can be evaluated and reviewed against all other beliefs in an agent's 'web of belief' (Quine & Ullian, 1970) in order to improve the agent's reliability overall—*belief holism*.

While weak foundationalism seems promising, Klein and Warfield (1994) make the startling claim that more coherent sets, even those formed reliably, are often less likely to be true than less coherent sets.

> What we wish to point out is that coherence, per se, is not truth conducive; that is, we will argue that by increasing the coherence of a set of beliefs, the new, more coherent set of beliefs is often less likely to be true than the original, less coherent set.

In essence, they argue that whenever an additional belief is added to a coherent set of beliefs, the likelihood that a belief is false increases. As each belief is added to a coherent whole, the likelihood that a belief is false increases. Holism means that a maximally coherent set of beliefs is almost certain to contain falsehoods—thus negating the purported benefits of coherence.

I overcome this objection by declaring holism to be false. Beliefs are not justified by cohering simultaneously with a whole system nor are beliefs justified locally. Instead, coherence adds justification choristically. I use the term 'chorism' (see §1.2) to explain how *limited* sets of mutually supporting apt beliefs improve performance success in a context C, towards goal G. At any time t_n, an agent has sets of justified cohering beliefs that may be inconsistent with other sets of beliefs. Sets of beliefs are revised and coordinated by the agent in turn, and it is this coordination of cycling sets of beliefs that best explains epistemic success. Achieving performance success requires not only coherence, but also distinction between sets of beliefs, and it is the ebb and flow of coherence and distinctness that leads to homeostatic epistemology (see Chapter 4).

In this chapter, I begin (in §3.2) by summarizing the argument for weak foundationalism, specifying how beliefs with low initial degrees of warrant contribute to an agent's success, and outlining some principles of coherence. In §3.3, I discuss holism, including cross-level coherence. In §3.4, I provide an explanation of a Bayesian coherentist account. In §3.5, I present the concern that coherence does not lead to truth and respond to this concern (in §3.6) before presenting an argument for coherence as maximal fittedness in §3.7. In §3.8, I advocate the value of distinction as a third source of epistemic justification, and in §3.9 I defend justification as the coordination between cohering and disintegrating sets of beliefs.

3.2   WEAK FOUNDATIONALISM

Foundationalism attempts to "axiomatize" knowledge, then bootstrap from basic beliefs to complex ideas—see Sosa's (1980) account of the 'pyramid' of foundationalist justification. Both empiricists and rationalists have tried to justify beliefs with foundationalist approaches—consider Hume's program of impressions that iteratively build to abstract ideas, or Descartes justification of beliefs generated from first principles that were themselves derived from self-evident truths—but both have failed to explain how normal, everyday beliefs count as knowledge. Most beliefs do not arise from self-evident truths or fully-informative sensory information, which is one of the reasons coherence is valuable, as it is meant to shore-up beliefs generated through fallible or dubious processes such as perception and memory. Furthermore, it is a means of managing possibly inconsistent or incomplete data.

However, by itself coherence has a limited capacity to justify beliefs—indeed Schlick thought that relying on coherence was "an astounding error" (Sosa, 2009a, p. 82). A madman, for example, could have a coherent set of delusional beliefs that have little bearing on reality and this obvious fact is known as the *isolation objection* (Olsson, 2012), which suggests that an internally consistent system offers little insight to either truth or reality. Related to this objection is the *alternative systems objection,* which raises the possibility of two independently coherent yet incompatible theories. Avoiding the isolation and alternative systems objections has led some philosophers to a weak foundationalism.

Weak foundationalism assigns a special warrant to highly reliable beliefs and then applies coherentist principles to scaffold a theory of knowledge (BonJour, 1985; Lewis, 1946). BonJour calls foundational beliefs "cognitively spontaneous" and Lewis calls them "supposed facts asserted." Cognitively spontaneous facts are not acquired by inference, but rather by highly reliable faculties such as sense perception, memory and introspection (BonJour, 1985). Lewis' account is more moderate, however, in that he requires 'supposed facts asserted' to have some measure of warrant for them to be useful within a coherentist justification of belief.

One advantage of weak foundationalism is that coherence raises the justification for beliefs with very low degrees of warrant, but a disadvantage of this approach is that the weak foundationalism appears to result in very low levels of knowledge of the external world. Perhaps it can identify true properties of physical objects and such, but how is abstract knowledge justified? Even perceptual success in the world may not set apart coherent theories—e.g., both Diffident and Dauntless could navigate across the Atlantic Ocean with a sextant regardless of Diffident's belief in a geocentric solar system and Dauntless' heliocentric inclination. Even weak foundationalism may not be sufficient for higher order justification.

BonJour (1985) attempts to improve weak foundationalism by defining coherence beyond mere consistency, to the mutual inferability of beliefs in the system and relations between these beliefs. He suggests that five principles govern coherence (pp. 97-99):

1. A system of beliefs is coherent only if it is logically consistent.

2. A system of beliefs is coherent in proportion to its degree of probabilistic consistency.

3. The coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and increased in proportion to the number and strength of such connections.

4. The coherence of a system of beliefs is diminished to the extent to which it is divided into subsystems of beliefs which are relatively unconnected to each other by inferential connections.

5. The coherence of a system of beliefs is decreased in proportion to the presence of unexplained anomalies in the believed content of the system.

BonJour's notion of coherence is that the beliefs in question must cohere with *the individual's* other beliefs, and not simply with *some* externally justified system of beliefs. I take BonJour to be largely right about coherence, but specifically, I concur that a system of beliefs is minimally coherent iff it is logically and probabilistically consistent. The coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and it increases in proportion to the strength of such connections. Coherence is decreased, meanwhile, by the presence of unexplained anomalies in a system of beliefs, or by its division into inferentially unconnected subsystems.

As this chapter progresses, I will argue that coherence adds justification when cohering beliefs increase the explanatory value of a set of beliefs towards

some goal, not necessarily in proportion to the *number* of inferential connections between beliefs. In any circumstance, there is an optimal set of beliefs and inferential connections that improve outcomes for the agent. If the optimal theory is right, then adding beliefs may increase the likelihood of truth to a point and additional beliefs thereafter may decrease the likelihood of truth. The greatest number of inferential connections occurs in reflective knowledge, and a lower number of inferential connections would occur in animal knowledge.

3.3   Holism

> Hornblower's mind completed the solution of the problem of the effect of the rudder at the same time as his senses solved it empirically (Forester, 1950)

Holism offers a different approach to justification. According to this view, beliefs are equal members of a 'web of belief' (Quine & Ullian, 1970), more like a ship being rebuilt on an open sea (Neurath, 1983/1932) than a pyramid standing on its apex. As Sosa (1980) eloquently states:

> …our body of knowledge is a raft that floats free of any anchor or tie. Repairs must be made afloat, and though no part is untouchable, we must stand on some in order to replace or repair others….what justifies a belief is not that it be an infallible belief with an indubitable object, nor that it have been proved deductively on such a basis, but that it cohere with a comprehensive system of beliefs.

Holism allows justification to flow downwards from rational considerations and upwards from perceptual beliefs. In this way, holism brings together internalist and externalist preferences, and beliefs are *informationally integrated* (Fodor, 1983; Weiskopf, 2008). Informational integration stems from Fodor's notion (1983) of isotropy, where justification for a doxastic state may be gained or lost due to changes elsewhere in the subject's doxastic state. Informational integration based on isotropy requires existing beliefs (and desires, hopes, and so

on etc…) to be updated "in concert" to reflect the impacts of new beliefs (Weiskopf, 2008). Extending Neurath's metaphor, beliefs are not reviewed individually when attention is directed to them, but instead are simultaneously and implicitly re-considered when new beliefs are formed. As Weiskopf (2008) explains:

> Informational integration is a relation that sets of mental states bear to one another such that a change to one such set causes appropriate changes in other, relevantly related sets of states. To count as a belief a state has to be part of a system of states in which processes of integration and updating function to keep the subject's mental contents in epistemic equilibrium to some degree or other. The hard work of spelling out how informational integration works requires saying under what circumstances a set of states is relevantly related to another and what changes count as appropriate (p.5).

Weiskopf explains the process of epistemic equilibrium by discussing the impact of discovering that a married couple has divorced on one's beliefs about their addresses. Weiskopf comments that "integration is part of the everyday dynamics of belief" (p.6) and they can occur either implicitly or explicitly. He imagines that while rapid, implicit mechanisms operate below the level of conscious awareness, conscious processes are slower and more deliberative. Though these conceptions are not central to the concept of informational integration, they are supported by the decision-making literature (Kahneman, 2011) and in line with Sosa's (2007) distinction between animal and reflective knowledge.

Sosa refers to integration as 'cross-level coherence':

> Proper reflective knowledge will after all satisfy requirements of coherence, which means not just logical or probabilistic coherence of the respective belief contents, but also the mutual basing relations that can properly reflect such coherence among the contents. Cross-level coherence, from the object to the meta, and conversely, is a special case of such coherence, and it imports 'guidance' of the animal belief by the

relevant meta-beliefs (or in other words, basing of the former on the latter) (Sosa, 2011, p. 13 footnote)

What is clear from Sosa's description is that he agrees with the first two of BonJour's criteria: logical and probabilistic consistency. Cross-level coherence entails that animal beliefs and reflective beliefs agree, even though reflective beliefs have greater justification.

Take, for example, the sailor Dauntless on her ocean-faring yacht. She has an apt animal belief that the boom is about to swing across the boat. Her animal knowledge is built up from quite complex muscle memories from many experiences tacking across the wind. She has tactile memories of the way the wind changes its impressions upon her face as the boat swings dead into the wind, then grips orthogonally the other way. She knows when to duck her head. This animal knowledge is justified. Even more impressively, if asked about why she ducked her head at the precise moment she did, she can justify that decision with reflective knowledge of sailing theory. She can explain how the boat lists starboard on a port tack, how it stabilizes directly into the wind, and how the boom comes across *just* as the boat begins to list towards port and wind collects in its sails. Dauntless' reflective beliefs cohere with her animal knowledge.

Plantinga (1993) raise an objection to coherentism, however. He argues that a subject could go his whole life without managing to achieve logical consistency across his beliefs. If coherence is necessary for knowledge, then we could not possibly know anything. Conversely, if we do know, then some less onerous notion of coherence must be agreed upon. Plantinga argues that we must not aim

for possibility in the broadly logical sense, or even first-order logic, but that we instead should aim for a weak coherence:

> *Weak coherence*: impossibility that would be obvious after a certain degree of reflection.

Nonetheless, even weak coherence spells trouble for knowledge. It is unclear that humans have sufficient reflective capacity to always determine that beliefs are incoherent, yet they are still knowledgeable in many ways.

Another criticism stems from the lottery paradox (Gwin, 2011; Hawthorne, 2004a), where an individual might rationally believe the ticket they are holding will not win, and yet, simultaneously know that there is at least one ticket in the lottery that will win. If this person went through each purchase of all the tickets, they would be conflicted due to the conjunction of all their beliefs not leading to the reality that one ticket will indeed be the lucky ticket. No matter of reflection or 'weak coherence' can overcome this contradiction. The lottery paradox prompts a Bayesian analysis of belief where the probability that any particular ticket will win is very small and the probability that the ticket will lose is very high. The Bayesian argues that it is rational at each purchase to believe the objective probability of getting a winning ticket—it is not inconsistent to acknowledge uncertainty and take objective chance at each decision. If Bayesian means can help interpret the Lottery paradox, perhaps it can also inform the notion of coherence itself? To explore this question, I turn to a weak foundationalist Bayesian model of coherence.

3.4   BAYESIAN COHERENCE

Coherence between propositions promises to fix the vexing circumstance of prior probabilities for Bayesianism (de Finetti, 1980). Prior probabilities require the knower to either know the base rates of the phenomena being considered or be justified in assigning equal likelihoods to each hypothesis. One way to circumvent a lack of knowledge that is pertinent to evaluating a proposition is to ensure that the prior coheres with existing, justified beliefs. It is worth noting this is only a problem for subjectivists. Objectivist Bayesians assign priors based on objective chances or frequencies. Subjectivists aim for beliefs to cohere with probabilistic axioms to avoid possible Dutch book scenarios. This section examines how subjectivists treat coherence.

Ewing (1934) took a consistent set to be coherent if each element followed by logical deduction from the rest considered together. Consider the following set of propositions:

T_1 = a triangle is made of three lines

T_2 = the sum of interior angles equals 180 degrees

T_3 = all lines connect with the vertex of each other line.

Each element of this set is consistent with the other elements and, more importantly, each follows deductively when the others in the set are considered simultaneously. Although Ewing's definition captures a maximal sense of coherence, it fails to capture the intuitive (albeit weaker) sense of coherence used

in common circumstances. Olsson (2012) considers the following propositions, where

A = "John was at the crime scene at the time of the robbery"
B = "John owns a gun of the type used by the robber"
C = "John deposited a large sum of money in his bank account the next day"

Intuitively, these three propositions are coherent, even though individually they do not deductively follow from the remaining two. It does not follow that John was at the crime scene at the time of the robbery because he owns the same gun as the robber and he deposited large sums of money in this bank account the next day. B & C offer circumstantial evidence for A, but they do not provide deductive evidence for A. The same logical processes can be applied to all three propositions to show that Ewing's definition is too strict.

Lewis (1946) improved on Ewing's theory by arguing that propositions 'support' one another if they raise the probability that a proposition is true. Here, 'support' is used in a weak probabilistic sense, rather than a strict logical sense.

Weak probabilistic support: P supports Q if an only if the probability of Q is raised on the assumption that P is true.

Lewis' definition explains why we think circumstantial evidence ought to be brought together. He argued that congruence between evidence collected from independent and partially reliable witnesses ought to increase our confidence in a hypothesis. As he explains:

For any one of these reports, taken singly, the extent to which it confirms what is reported may be slight. And antecedently, the probability of what is reported may also be small. But congruence of the reports establishes a high probability of what they agree upon, by principles of probability determination which are familiar: on any other hypothesis than that of truth-telling, this agreement is highly unlikely; the story any one false witness might tell being one out of so very large a number of equally possible choices. (It is comparable to the improbability that successive drawings of one marble out of a very large number will each result in the one white

marble in the lot.) And the one hypothesis which itself is congruent with this agreement becomes thereby commensurably well established (p.246)

Lewis says that even moderately reliable witnesses contribute something evidential towards evaluating hypotheses so long as there is a non-zero prior probability that their report is right. Consider two competing hypotheses. There ought to be a rational way to determine how the evidence collected should be evaluated with respect to each hypothesis. For example, if there are two suspects in a murder, there would be much overlapping evidence (such as crime scene evidence) and eye-witness report evidence, but there would be background information on each suspect that would affect how likely we might consider the evidence surrounding the crime.

Bayesian approaches to coherence create a procedure for deciding between hypotheses. First, the Bayesian asks: How expected is the result? There may be historical precedents, personality traits or circumstantial factors that set up a base level of likelihood that each suspect was involved in the crime. Second, a Bayesian asks: How reliable is the evidence? Evidence stemming from highly reliable sources, such as DNA-matching experts, ought to be considered more seriously than less reliable sources (for example, an absent-minded bystander's eye-witness report). Third, a Bayesian treats coherent evidence more seriously than incongruent evidence. So, if the absent-minded bystander claims he remembers orange headphones on one suspect and another eye-witness remembers seeing an .mp3 player, the coherence of their testimony ought to raise the seriousness of their claims. Bayesians treat the third of these, coherence, as *ceteris paribus* claims (i.e., all else being equal, evidence that is coherent is

preferred over incoherent evidence). In other words, coherence is the least valuable of the three considerations above, and knowledge about the prior likelihood of an event and the reliability of the evidence is of the greatest value (e.g., highly reliable DNA-test results trump coherent evidence from less reliable sources).

In addition to the non-zero prior, Lewis' theory requires both a positive and a negative thesis: The positive thesis is that coherence increases the posterior probability that x occurred with the number of consistent beliefs. The negative thesis is that the coherence of independent items of evidence has no impact on the probability of a conclusion unless each item has some credibility of its own. The outcome is that the degree of congruence is inversely related to the prior probability of the supported hypothesis, so agreement on something antecedently improbable gives rise to a high degree of congruence and should lead to notable belief revision.

There has been a response to this account regarding whether a prior probability must be greater than 0, which some commentators saying 'no' (BonJour, 1985; Siebel & Wolff, 2008) and others saying 'yes,' with some restrictions (Roche, 2010; Van Cleve, 2011). BonJour (1985) argues that "[w]hat Lewis does not see, however, is that his own [witness] example shows quite convincingly that no antecedent degree of warrant or credibility is required" (p.148). BonJour supposes that the independence of witnesses, so long as their accounts cohere, is sufficient to "eventually dictate the hypothesis of truth telling as the only available explanation of their agreement" (p.148).

The intuition BonJour is alluding to here is the unlikelihood of two independent witnesses coming up with cohering propositions without good reason—the most likely reason being that they both saw the same thing and are reporting on it. How could two people easily imagine the complex details of an idiosyncratic event? The worry, as Bovens and Hartmann (2004) point out, is that people can imagine cohering ideas because they are the most obvious, rather than because they have specific knowledge relating to an event. For example if one asks a westerner to think of a fruit, they often say 'banana' because it is associatively connected (see Ch.4 Kahneman, 2011). There are typical or socially acceptable suggestions that may attract credit and status for the individual, but these may have been generated through sheer guesswork, and may have no bearing on what actually occurred.

Below, I outline Olsson's (2012) argument that some antecedent degree of warrant must exist for coherence to bring justification. Consider the following two testimonies.

Let E_1 be the proposition that the first witness reports that A, and let E_2 be the proposition that the second witness reports that A. Consider the following conditions:

Conditional Independence

$$P(E_2|E_1,A) = P(E_2|A)$$

$$P(E_2|E_1, \sim A) = P(E_2|\sim A)$$

Nonfoundationalism

$$P(A|\ E_1) = P(A)$$

$$P(A|\ E_2) = P(A)$$

Coherence Justification

$$P(A|E1, E_2) > P(A)$$

The conditional independence requirement means that there would be nothing connecting testimonies apart from seeing the actual event in question, thus ensuring no influence or communication between the parties, and no third party interference to adulterate the probabilities. Nonfoundationalism states that by itself, neither testimony offers insight on the probability of A (replicating the circumstances where neither witness' testimonies have any prior probability). The coherence justification claims that when considered together, coherence between two reports does provide justification, and therefore does increase the likelihood of A. From conditional independence and nonfoundationalism it follows that:

$$P(A|E1, E_2) = P(A)$$

Thus, as Olsson (2012) says, "combining collectively independent but individually useless testimonies, however coherent, fails to give rise to anything useful." Note the difference between nonfoundationalism and weak foundationalism.

Weak foundationalism

$$P(A|\ E_1) > P(A)$$

$$P(A|\ E_2) > P(A)$$

Weak foundationalism with conditional independence implies coherence justification. The combination of cohering testimonies increases the likelihood of *A* more than a single testimony, and in this way, coherence can boost credibility or justification even if it cannot do so from scratch (Van Cleve, 2011). Employing apparatus from foundationalism is one way to save coherence from the *isolation objection*—the worry about how purely internal relations could specify the conditions of an external reality.

There are other characterizations of independence and nonfoundationalism (e.g., Huemer, 2011), although their power to explain various intuitions has been questioned (Olsson, 2010; Scheibehenne, et al., 2012). Contrary to Huemer's view, the value of conditional independence has been shown in the success of Bayesian networks (Pearl, 1985, 2000) and standard applications of probability theory (Scheibehenne, et al., 2012). Even BonJour (1999) has retracted his earlier coherentist position.

3.5 COHERENCE DOES NOT LEAD TO TRUTH

If some variant of weak foundationalism is indeed the way to save coherentism, what follows now? The idea might be that ideal epistemic agents ought to constantly improve the coherence of their beliefs, and that coherence— with reliable foundations—will eventually lead them to truth. However, Peter Klein and Ted Warfield (1994) (henceforth KW) suggest a disturbing situation where, contrary to expectations, a *higher* degree of coherence can in fact *lower* the probability of a set of beliefs. KW's counterintuitive result comes from the following argument from their paper:

In order to show that a coherentist must prefer coherence to truth, only two premises are needed:

(P1) Any consistent set of beliefs, B, is more likely to be true than any set, B*, which contains all members of B and at least one additional belief, so long as at least one additional belief in B* has neither an objective probability of 1 nor is entailed by B.

(P2) One strategy for converting a less coherent set of beliefs into a more coherent set of beliefs is to add a belief (to the less coherent set) which has neither an objective probability of 1 nor is entailed by the less coherent set of beliefs.

(C) a more coherent set of beliefs resulting from the addition of a belief to a less coherent set of beliefs is less likely to be true than the less coherent set of beliefs (Klein & Warfield, 1994, p. 130).

KW defend P2 by claiming that a set of beliefs can be rendered more coherent in two basic ways:

(a) the Subtraction Strategy in which a belief (and perhaps with it many more) is subtracted from a less coherent set, thereby rendering it more coherent;
(b) the Addition Strategy in which one or more beliefs are added to a consistent set of beliefs to render the set more coherent (Klein & Warfield, 1994, p. 130).

They also acknowledge a third strategy that employs both (a) and (b).

I take issue with P1 and P2, which leads me to reject C as a consequent of P1 & P2. However, there is much clarity to be gained from discussing both the argument and my own defense of coherence that emerges from it.

First I wish to address the commitments of P1. Is it necessarily true that any consistent set of beliefs smaller than another consistent set of beliefs is more likely to be true, where a belief added to the larger set is not entailed by the existing beliefs and objective probability $\neq 1$? Suppose I am reliably looking at my hand and I come to believe I have a hand. Is it any less likely to be true if I also form beliefs that I have a thumb, forefinger, a wrist, and so on based on the same perceptual experience? One can logically have a hand with no fingers, so having fingers is not entailed by seeing a hand. Perhaps KW would reject my

example because they expect new beliefs in a set to have to been generated via independent processes? Supposing that is true let us imagine that I believe I have a hand because I have a perceptual experience of my hand and I have a memory that I saw my fingers yesterday. These perceptual and memory beliefs were independently generated,[54] are consistent, and they form a larger set. Is it less likely to be true that I have a hand and that I saw my fingers yesterday, than the likelihood that I have a hand?

Ceteris paribus, the more beliefs one has, the more likely they were generated in ways that make them false. But, for any *particular* set of beliefs, it is not obvious that the larger set must be less likely than the smaller set when all relevant information is considered. Suppose I am just living in the normal, mundane world, with perceptual processes and memory processes operating reliably in normal conditions. In this context, the larger set of beliefs about my hands and fingers could be just as likely to be true as the belief that I have a hand. That is, there are plenty of exceptions that make P1 false. Nonetheless, the *ceteris paribus* intuition does have some value. We may not know, for example, whether we are in good conditions or bad conditions, which means that perhaps we ought to assume that larger sets of beliefs are more likely to be false than smaller sets of beliefs? For now, I put this issue aside and turn instead to the question of whether KW notion of coherence in P2 is correct.

---

[54] One might argue that the memory of seeing my fingers yesterday is not independent because it is foundationally justified by a perceptual experience akin to the one I am now having. However, I will treat the two perceiving instances as independent due to differences in context, where some aberrant factor may have influenced my judgment (e.g., perhaps I only dreamt that I saw my fingers yesterday and I have misremembered this dream as a veridical perception).

In P2, KW claim that coherence can be increased through the addition of a logically consistent belief that does not have an objective probability of 1 and is not entailed by the less coherent set of beliefs. I want to separate the notion that coherence is simply a synonym for 'logically consistent' from the more nuanced notions introduced by BonJour (1985) and Lewis (1946). BonJour pointed out that coherence involves logical consistency—the strength and number of inferential *and* explanatory connections between beliefs. (Notice that KW explicitly remove entailment from P1 and P2—a point that I get back to in §3.6). I will refer to these more nuanced ideas about coherence as I proceed through this discussion, but the first explanatory connections relevant to coherence that I consider are informativeness and explanatory relevance.

Coherence as a source of justification is not merely a matter of the number of logically consistent beliefs, but also the degree to which they are informative to each other—and by 'informative', I mean in Shannon's (1948) sense of increasing the probability of the other beliefs in the set by being true. Consider the hypothesis BD and the radio report RR:

BD: The bushfire will destroy Farmer Bob's barn.

RR: The bushfire has shifted its path westwards.

Hearing a radio report RR may be evidence in support of BD. However, other reports—such as internet reports of the destruction of houses along the fire's path or the heat of the fire— do not bear on BD, despite being logically consistent with BD and RR,. In fact, there are an infinite number of beliefs that are logically consistent with BD but have no relevance to evaluating BD. The beliefs that

increase coherence amongst beliefs—as a source of justification—are those that increase the likelihood that the hypothesis is correct, but KW do not take informativeness or relevance into consideration in P2.

Critically, relevance is not just informative relative to the hypothesis, because if epistemology is an agent-centered endeavor (as virtue epistemology and Bayesian epistemology recommend), beliefs relevant to the hypothesis must also relevant to the goals, capacities and context of the agent. While the firefighter's goal is to prioritize limiting the loss of life, the farmer's goal is to minimize the destruction of his barn. Each individual has a domain of competence that will affect the justificatory value of a belief to the coherence of a set, and the firefighter's belief BD and the farmer's belief BD may both be justified within different integrated sets of beliefs. The firefighter's belief BD might be inferred from his beliefs, RR, FD and BFD.

FD: the bushfire will destroy all properties west of Flowerdale

BFD: Farmer Bob's barn is west of Flowerdale

On the other hand, Farmer Bob's belief BD is justified perceptually and through testimony (RR). Both Farmer Bob and the fireman believe the same radio report RR that the fire has shifted its path, but they use that information differently to justify their belief BD. In sum, the value of coherence as a source of justification depends on whether when a belief is informative and relevant to the goals and context of an agent, rather than whether it is logically consistent with other beliefs.

Olsson (2012) recasts KW's argument from the perspective of information:

P1: Coherence can be raised by adding more information that explains the information already in the set.
P2: The more information in the set, the more likely some elements are untrue.
C: Coherence is not truth-conducive.

Olsson (2012) illustrates this argument in the following way. Suppose there are two sources of information about Tweety. Jane says that Tweety is a bird and Carl says that Tweety cannot fly. The resulting set S is not particularly coherent. After all, typical birds can fly, so it would be odd if Tweety was a bird and could not fly. However, suppose a third source presents a third piece of evidence about Tweety, informs us that Tweety is a penguin. This new set, S', consists of three statements and seems more coherent than the first set. It is explanatory, and more information has increased coherence. KW argue that the likelihood of all three statements being true is less than the likelihood of two statements being true, which means that the conjunction of fewer propositions is more probable than the conjunction of more propositions. As a result, more coherence does not imply a higher likelihood of truth in the sense of higher joint probability.

Olsson also argues that the joint probability of a set of reported propositions is as follows.

A_1 = "Tweety is a bird"

A_2 = "Tweety cannot fly"

A_3 = "Tweety is a penguin"

The less coherent set of S consists of A_1 and A_2, whereas the more coherent set S' consists of A_1, A_2, and A_3. Intuitively, if C denotes the degree of coherence, then:

$C(A_1,A_2) < C(A_1,A_2,A_3)$.

Yet, given the increased informational content of S', it is less likely than the smaller set S:

$P(A_1,A_2) > P(A_1,A_2,A_3)$.

This formulation treats the propositions as facts, instead of what they are: testimonies. Olsson reformulates this on the basis of testimonial evidence. Keep in mind that reports from different witnesses might have different reliabilities that might affect the outcome.

E_1 = "Jane reports that Tweety is a bird"

E_2 = "Carl reports that Tweety cannot fly"

E_3 = "Rick reports that Tweety is a Penguin"

Using the *principle of total evidence,* [55] the true probabilities are:

S is $P(A_1,A_2 \mid E_1, E_2)$

S' is $P(A_1,A_2,A_3 \mid E_1, E_2, E_3)$.

However, even when the problem is restated in these terms, smaller sets are more likely to be true than bigger sets (2012).[56]

$P(A_1,A_2 \mid E_1, E_2) > P(A_1,A_2,A_3 \mid E_1, E_2, E_3)$.

So, have KW and Olsson shown that increased coherence reduces the likelihood that a set of beliefs is true? I believe that the problem is not a problem

---

[55] One of the features of Bayesian confirmation theory is that "a rational believer will proportion her confidence in a hypothesis H to *her total evidence* for H, so that her subjective probability for H reflects the overall balance of her reasons for or against its truth" (Spiegel, 2012).

[56] Even if the restatement does not resolve K&W point, Olsson's reformulation suggests that all evidence must be taken into equally account (beliefs, testimonies etc…). Coherence operates not just on testimony, but on testimonies that have been reflectively considered in light of other evidence.

of coherence, but rather is a problem coherence as it has been formally constructed by KW and Olsson (2012).

Let us return to Olsson's example of Tweety. The purpose of S and S' is not to form a set of consistent true beliefs, but rather to determine what Tweety is.[57] Tweety turns out to be a penguin, which means that the only belief necessary for the maximum coherence of the set, qua explanation, is E_3 (assuming E_3 is reliably formed). The fact that E_1 and E_2 exist and are logically consistent with E_3 does not add much justification to E_3. To demonstrate this, I use KW's subtraction strategy P2 (a). Examining the set $[E_1, E_2, E_3]$, we should subtract E_1 and E_2, because they are entailed by E_3. That is, the fact that Tweety is a penguin entails that Tweety cannot fly and that Tweety is a bird. We would only keep E_x in a set if it added explanatory information, e.g., suppose additional evidence E_4 exists.

E_4 = "Sarah reports that Tweety is 60 million years old"

Adding E_4 may add information, but it does not affect the coherence of the existing beliefs because it does not change the credence of any of the beliefs. However, if Tweety turns out to be a Waimanu penguin—an ancestral penguin that is classified as a bird and probably also capable of flight (Slack et al., 2006)—then E_2, ('Tweety cannot fly'), may be explanatorily inconsistent (though logically consistent), and therefore should be subtracted from the set. One of the problems Olsson appears to have is the problem faced by classical conditionalization that no learning episode can be undone by further applications of the same rule (see §1.7.3). The way humans justify beliefs is by choosing those

---

[57] KW have a similar purpose in their example of a crime scene, they wish to identify a murderer.

most explanatory, regardless of the order in which information is learned. The point here is that the normative value of coherence to the truth of a set of beliefs is more than logical consistency or even informativeness but explanation, more precisely, explanation relative to some goal. Adding a belief to a set raises the probability of another belief in the set when it plays the appropriate role, which I define it in the next section.

I want to end this section by noting KW and Olsson's excellent observation—that increasing beliefs in a set increases the odds that some of them are false—is a problem fallible belief-producing processes generally, not a problem for coherence per se. I take the threat of increasing beliefs seriously enough to suggest in addition to reconsidering coherence, we must also redefine the justification of beliefs relative to limited *choristic* sets, rather than beliefs considered holistically.

3.6   WHAT IS COHERENCE?

In P2, KW define coherence as consistency *separate to implication*, but removing implication actually side-steps what coherence is. Coherence adds strength to independently and reliably informed beliefs *in virtue* of its implications (that is, its relationships and explanatory connections to other beliefs in a set). Beliefs that add coherence add logical glue to the existing beliefs in the set. Some beliefs, because they fail to add understanding or depth, may not increase these relations, and therefore not increase coherence. Coherence is *all* about the relations between beliefs, and not necessarily (or even generally) deductive relations, but also inductive relations (e.g., I believe the swan I just saw

is white and the swan I'm seeing now is white) and abductive relations (e.g., the most likely explanation for beliefs p_1, p_2 and p_3 is that my car was towed away). However, coherence goes even further than these relations; beliefs cohere with other beliefs due to causation, contiguity and natural laws, as well as semantic, temporal, or conceptual associations and processes.

In trying to make sense of these varied connections, I am reminded of Lewis Carroll's (1895) parable, of *What the Tortoise said to Achilles*. In the parable, the obtuse tortoise points out to an exasperated Achilles that the paradox of logic is that you always seem to need to add further premises in order to make the jump from a set of premises to conclusion. This prompts the question, is there ever enough coherence between our beliefs to adequately justify them? Quine & Ullian(1970) addressed aspects of this dilemma by situating fundamental beliefs more centrally in his web of belief so that there was some order within the whole, but, explicating precisely the relationship between fundamental beliefs and peripheral beliefs is yet to be agreed on or formalized. Trying to understand the relations between beliefs and how they entail or connect with one another—and when any finite set of beliefs is enough for some goal in a particular context— remains the central puzzle of coherence.

One distinction that that should be made explicit is the difference between coherence as a *description* of relations between beliefs and coherence as a *normative* notion that plays a justificatory role in raising the probability that beliefs are true[58]. Beliefs may cohere with one another poorly, such as when a

---

[58] Thanks to Michael Devitt for making this distinction clear to me (personal communication, August 18, 2013).

person justifies their drunk driving by explaining that they have been drunk driving for 20 years and have not yet had an accident, and argues that they are therefore unlikely to have an accident in the future (a misuse of induction based on confusing his extraordinary luck with application of skill). For the sake of the rest of the discussion, please assume that when I mention coherence, I refer to normative coherence (i.e. that coherence raises the likelihood that a belief is true when it arises as a result of an appropriate application of normative rules of inference).

I consider a coherent set of beliefs to be a set that is *satisfactorily* held together. By removing implication, KW's notion of coherence is equivalent to stacking a load of matching planks next to one another, eschewing fastenings, and complaining that there is no house, only planks.[59] Of course no house exists—a house is built when planks are put together in an emergent way that produces an object greater than its parts. The relationships between beliefs are the nails, glue, and joinery that form greater knowledge. Coherence is the joining together of beliefs. To be fair, some beliefs can be the joinery and some behave as planks, and sometimes beliefs butt against one another weakly. Some beliefs (to stretch the analogy) like a bridle joint, add strength because they have overlapping implications, while some are superfluous in many contexts, but are needed under great stress (perhaps a virtue of reflective rather than animal knowledge?). Other beliefs are part of a contingency design, which will ensure that the roof remains even when some planks or other joinery fail or rot. Beliefs that yield coherence

---

[59] And the more planks one stacks, the more likely some of the planks will be termite-ridden or suffering from fungal rot.

can be hidden premises that ensure entailment, and by making hidden premises explicit reflective justification improves. It is also possible that non-beliefs are the fastenings. Perhaps the rules of logic and inference, like the rules of grammar, are not explicitly represented in the agent's cognitive architecture (Devitt, 2006)? In truth, we do not yet know the relations that precisely define or justify coherence, but we know that they must exist, otherwise there could be no house. Is there a way forward with coherence that limits the danger of endlessly increasing the number of propositions, and therefore increasing the odds of falsehoods?

A cohering belief that makes a set of beliefs more likely to be true must add logical glue and relational fastenings—i.e. explanatory clarity. Sets of unrelated but consistent beliefs are not coherent, and they add danger (increase likely falsehood), without adding veritistic benefit. Ironically, KW hint that they understand this in P2 (a) and (b), where they note that *either* adding or subtracting beliefs can bring more coherence. This suggests that KW appreciate that normative coherence represents more than logical consistency; instead, coherence reaches some maximal point for some set of beliefs for some goal. So what makes someone remove a belief to make something more coherent? Removing or adding a belief is based on how contributive to the truth the belief is.

I argue that there might be a maximally coherent set of beliefs for which no further propositions would be informationally or explanatorily useful to justifying a belief for the agent in a context towards a particular goal. Limiting the number of beliefs in a set is one method for reducing the threat of falsehood through ever-increasing propositions. Note that this is contrary to the typical Bayesian

solution—that it is always better to add more informative propositions to increase the probability that a hypothesis is true. As a normative guideline, the Bayesian answer to coherence leads to an infinite search for more information relevant to a hypothesis. By adopting ideas from virtue epistemology, however, I presume that meeting a threshold for cohering beliefs can suffice for knowledge and this offers a way to limit how many propositions must cohere.

3.7   COHERENCE AS MAXIMAL FITTEDNESS

Why do KW and Olsson—and holism more generally—suppose that increasing the number of logically consistent propositions ought to increase justification? I argue that this intuition is not about justification, but rather is about detail or comprehensiveness. Our ideas do not necessarily become more coherent because we have more of them. Instead, I argue that coherence is a measure of how well propositions fit together to serve some goal in a particular context. Suppose there was a maximum fittedness. Just as there is a perfect cohesion between carved pieces of green oak to make a wood frame for a home, there is a perfect level of cohesion between any number of propositions to serve some purpose of the agent (see *Figure 3.1*). There is nothing to be gained by adding more pieces of wood to the frame once the ideal construction has been built. The frame's job is to be sturdy and mutually supportive. Once this has been achieved, its job is done. The best solution is the most efficient and effective.

> *Efficient*: Requires the least wood (cheap and environmentally sensitive) and the wood chosen lasts the longest (and grows harder and becomes more sturdy over time).
> *Effective*: Can hold up the roof, is suitable for long-term weather conditions, allows the greatest distance between supports for windows, doors and so forth.

Analogously, one ought not clutter or clog the mind with inefficient or ineffective facts. For example, having true beliefs about how many grains of sand exist on the beach is pointless (Pritchard & Turri, 2012). Conceiving knowledge as just objective facts calls its value is in question, but, if knowledge is defined as a) relative to an agent and b) relative to an agent's goals, then knowledge is not clutter. The intuitive pull of the 'grains of sand' example is the pointlessness of such facts to actually help anyone achieve anything. The value of beliefs is that they progress a creature to some end.[60] As such, coherence as a source of justification should not aimless consistency or an infinite number of connections between beliefs, but rather should offer connections that allow the agent to achieve success. After all, the propositions that define a triangle are mutually supporting and do not need any more propositions—see Ewing's (1934) original definition.

---

[60] It is not relevant here to delve into the details of the value of the agent's goals. Perhaps there are objectively valuable actions—such as making moral decisions, or making decisions to improve one's health or happiness or the health and happiness of others?—Or are there subjectively valuable actions such as putting in the least effort, being comfortable or getting a thrill? Regardless, the answer to this has little bearing on my point.
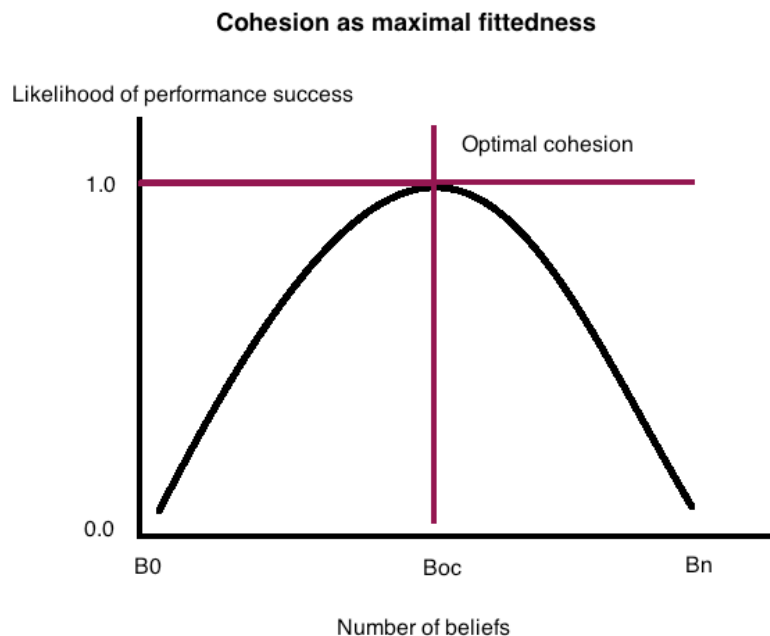
Cohesion as maximal fittedness

Likelihood of performance success

Figure 3.1 Cohesion as maximal fittedness. A set of beliefs (Boc) is maximally cohesive when it reaches the maximal likelihood for performance success for an organism in a particular circumstance. The optimal cohesion balances the benefits of beliefs adding explanatory strength to existing beliefs with the risk of adding too many irrelevant beliefs.

What are the consequences of coherence if it is defined as maximal fittedness, and if, for any set of propositions, coherence increases up until they reach an optimal point. Less coherent sets are those that do not manage to serve their function ideally—like a green wood frame missing an arch. There would be some coherence, some success achieved by the construction, but it would be at some risk—slightly unsafe to use Sosa's (1999) terminology. Note that not just any additional wood (proposition) would do; only the missing pieces add coherence. Similarly, a poorly coherent set of propositions would not be made more coherent just by adding more propositions that add little additional strength to the case.

Consider the following example: A single security camera that captures the thief's face on film is useful to identify a suspect. Once this footage is obtained, no further security cameras are necessary to prove the thief's identity. The footage of many security cameras may increase the total information, but it does not improve coherence unless the additional cameras do more than place the thief at the crime scene (e.g., they show the thief stealing a particular object). Propositions to increase coherence must be shaped finely to add to the mutual supportiveness of the set, but (unlike Ewing's stringent demands on mutual coherence) no one proposition must be useful to all other propositions. Instead, each proposition must have a relationship with some subset of propositions to build a strong component of the entire structure.

A skeptic might wonder if there is such a thing as a single optimal set of beliefs. What if there are many different optimal or close-to-optimal solutions that work in different ways? For example, there are many near-optimal ways of playing chess, and each grandmaster has their optimal playing style. In response to this criticism, it is worth pointing out that while there are differences between these playing styles, grandmasters share many of the same strategies and techniques. There are books of agreed lore on the strategy and tactics of chess that underpin much of the grandmaster's success; for example, principles such as 'control the center', 'develop the pieces as rapidly as possible' or even the simple 'avoid having your pieces taken.' (Oaksford & Chater, 2007, p. 283). Similarly, there are some agreed rules of inference (Oaksford & Chater, 2007) (*mondus ponens*, *modus tollens*, *biconditional introduction* etc.) and explanation (Hempel,

1965; Salmon, 1971, 1984) (e.g. *deductive-nomological model, statistical relevance, causal mechanical* etc.) to improve coherence (as a source of justification). My view does not depend on a single set or arrangement of beliefs necessarily and sufficiently described. Optimality is measured in terms of success for the agent, and optimal means that a person has balanced finite resources to resolve their current predicament. Did the grandmaster win the game? Did they win their game faster than another person or retain the most pieces at its conclusion?

A skeptic might also point out that most of our belief sets are not optimally cohesive, but rather are satisficing (see *Figure 3.2*). We are justified because we get enough of what we need to get by, and striving for optimality does not mean that we fail if we 'just get by.' Sosa (2009a) supports the idea of 'sufficiency' with regards to coherence. He states:

> It is very hard to see how to draw a line above which lie the degrees of comprehensiveness and coherence that suffice for knowledge…compare a concept like that of being tall. That is presumably to be defined in some such way as this: being sufficiently taller than the average. Presumably someone just infinitesimally taller than the average is not tall. One has to be taller than the average by some margin, one has to be 'sufficiently' taller than the average. But how do we define that margin? Is there, even in principle, some way to capture our actual concept of tallness by means of some such definition? There seems no way. Yet we do surely have and use a concept of tallness, do we not? Why can't we view epistemic justification similarly in terms of "sufficient" comprehensiveness and coherence? (Footnote, p.150)

Sosa supposes that epistemic justification is a matter of finding sufficient coherence above a certain threshold, and it seems right that satisficing coherence meets a minimal level to ensure performance success—see Godfrey-Smith (2001) for a discussion of satisficing as it pertains to evolutionary adaptation.
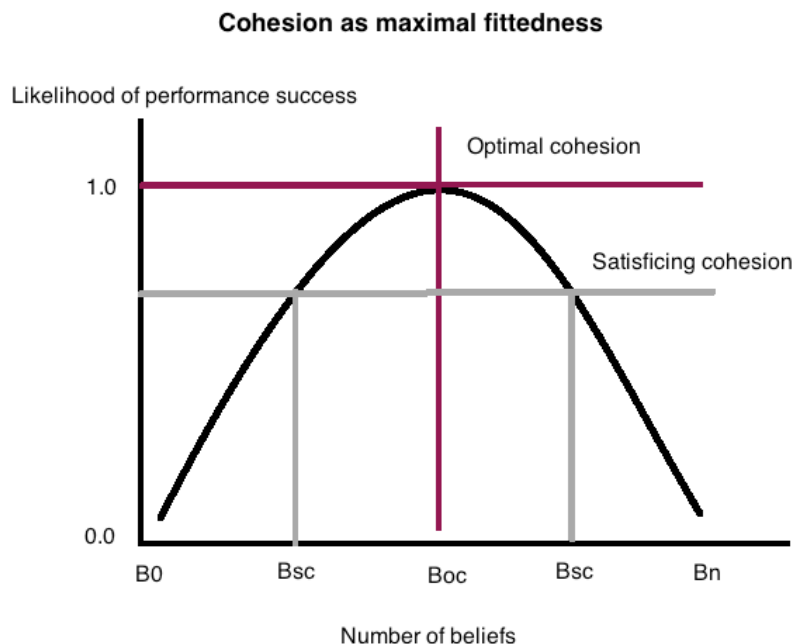
Cohesion as maximal fittedness

Figure 3.2 Cohesion as maximal fittedness plus satisficing cohesion conditions that meet a threshold for performance success.

One consequence of the view that coherence is maximal fittedness might be that removing one proposition would have disastrous consequences on the whole, rendering it suboptimal. Certainly, such a rigid structure would never work in the flurry of uncertain reasoning. It must be kept in mind that a green wood frame is designed to hold up a home. It does not necessarily require a particular design: A frame might hold up the roof with fewer propositions using different shapes; it might hold up the roof with existing propositions arranged differently; and it may hold up the roof with a few more propositions. Similarly, the mind does not depend on only one set of beliefs by which to glum onto the truth, and it does not depend only on sight or only on touch. Sets of beliefs can be rearranged (imagined, considered, deliberated on) to test different structures. Rearranging sets is the process of hypothesis testing—of finding what structures are possible

given the number of beams, just like a game of trying to build 3-D shapes from a set of match sticks. The main point is that coherence does not increase along with the number of consistent propositions. Instead, coherence hits a threshold when a set of propositions enables the agent to perform optimally (although most of the time agents are in a process of flux with their beliefs, attuning them and revising them in response to, or in anticipation of, their changing environment).

Coherence means more than just 'mutually supportive', or 'consistent'. Coherence requires an ideal number of propositions to achieve some goal, and this is where coherence and consistency differ. While coherence is relevant to an agent, consistency is only relative to belief (objectively measured). My approach treats coherence as a *limited* interdependency that supposes that a *limited* set of beliefs functions better than the individual beliefs would if they were treated separately. This is not the same as encapsulation, because beliefs are mostly not locked away from other beliefs.[61] Consider the mutual support of green oak timber planks for a house: There are a set number of planks required for a particular design no more (and no less) are needed for ideal coherence.

3.8   THE VALUE OF DISTINCTNESS

In this section, I justify epistemological distinctness between beliefs.[62] The value of coherence lies in its role as part of a two-stage process of belief revision. Sometimes incoherence (between subsets of our beliefs) is necessary and it prompts us to 'go back to square 1', 'reset' thinks, or 'shake things up.' We reach a dead end, where we have achieved great coherence, but a not-so-great design. It

---

[61] Beliefs can work in functional, computationally limited sets, yet still be cognitively accessible to other sets of beliefs. Unlike cognitively impenetrable modules described in the *Modularity of Mind* (Fodor, 1983).
[62] Separate to the metaphysical notion of distinctness in Descartes (Hoffman, 2002)

is better at these points to disassemble the set into its constituent parts and reconceptualize what we want. Slavish devotion to coherence can leave us vulnerable to changing contexts. To use the analogy of the sabre tooth cat, the canines get longer and longer in order to target specific prey (goals), but if the prey become extinct (context change), the cat will become extinct too because they cannot adapt to the new circumstances (Van Valkenburgh, 2007). To be adaptive, coherence must be balanced with disintegration or distinction.

Disintegrative processes ought to be implemented when our systems begin to be sluggish—like a computer that needs its hard drive defragmented. Notice, for example, the value put on 'analytic work' in analytic philosophy—insights emerge from isolating and disentangling ideas from one another. Taking beliefs out of context allows the philosopher to focus on them without the influence of other ideas. Consider also the value in 'unbiased' observation,[63] where putting aside preconceptions and expectations makes us realize that we sometimes do need to overhaul our approach to ideas to find the truth. A relentless focus on coherence will not lead to truth. Useful philosophical work occurs within a set of assumptions, relations, and outcomes that can then slot into other sets of beliefs, but limited sets can be tinkered with without sinking the ship of reason.

Consider the concept of a 'fast deck' in trading card games. Trading card games require players to build strategic decks of cards from hundreds of possible cards and then play against one another. If a player selects a fast deck, she has selected cards that will help her to move efficiently through them for maximum

---

[63] Even though totally unbiased observation may be impossible (Chalmers, 1989)

effectiveness (e.g., awarding more actions such as attacks or buys). However, selecting a fast deck requires a great deal of skill, and novice players are revealed by their slow or 'flabby' decks, which they have selected because they overvalued expensive power cards and underappreciated the value of cards that work with other cards to produce larger effects. A flabby deck stops cards circulating quickly, which prevents double or triple plays or synergistic power-ups. Flabby decks depend on too many rounds to wash through the cards. A deck with the wrong cards or too many cards fails to get the player what they need, when they need it. Sets of beliefs, like decks of cards, should be fast. Beliefs work best in efficient, collaborative groups, not added merely to have more. The risk of 'flabby beliefs' is real in memory research, where remembering too much from one's past can prevent one from remembering pertinent facts today (Nairne & Pandeirada, 2008), or being able to make distinctions (Luria, 1987).

Work that involves discrimination and distinction is important, and it is necessarily separate from synergies and interdependencies found in synthesis. In fact, most of the time, we believe little enclaves of propositions to a certain degree, despite an awareness that they do not perfectly cohere with the entire body of knowledge we rely on. Of course, there is nothing new in these ideas. Thomas Kuhn famously argued that the progression of science depends on periods of revolution rather than normal science. Given what we now think it is plausible to think of children as young scientists, testing hypotheses, it is reasonable to suppose that children (and adults) go through doxastic periods of stability and disintegration. If coherentism means to converge, connect, and

decipher a single message from a set of data,[64] then multiple-hypothesis-holding means to diverge, separate, and elucidate many interpretations from a set of data,[65] and both convergence and divergence have their place.

Distinctness is a source of justification when we can ascertain no necessary, causal, or explanatory connection between one set of beliefs and another. Rationalists might defend distinctness by alluding to the clarity of understanding or declaring the analytic truth of a set of beliefs. A virtue-Bayesian, however, justifies distinctness when a set of accurate beliefs (produced by adroit processes) is sufficient for a particular agent to perform successfully in a particular context. Take my belief that it is snowing outside and my belief that a diner has smoked meat sandwiches. My belief that it is snowing is part of a larger set of beliefs about weather and visual perception. My belief that a diner has smoked meat sandwiches is part of a set of beliefs about common menu items in Montreal diners. These sets of beliefs are valuable distinct from one another so long as no pertinent facts from one set affect the other. No menu information from the latter set will affect the weather. However, it may be that the diner does not have smoked meat sandwiches today because it is snowing and the delivery van could not deliver. Here the agent still knows that it is snowing, but now fails to know that there are smoked meat sandwiches at the diner. This failure of knowledge is easily explained by the missing belief connecting weather conditions and smoked meat delivery, but the newly integrated set of beliefs leads the agent to different

---

[64] Convergent thinking involves linking together three separate ideas through one unifying idea, (for example, noting that food, forward, and break can all be united with the word 'forward') (Guilford, 1967).
[65] Divergent thinking involves coming up with multiple explanations, such as the multiple uses for a brick (Schulkin, 2003, p. 4).

sorts of knowledge as their beliefs are updated. Of course, many beliefs remain distinct and irrelevant (for example, knowledge of my niece's birthday in six months time). A distinct credence (i.e. treating a proposition as probable that does not cohere with other, existing beliefs/propositions) remains valuable in so far as it obeys Bayesian norms in being valued proportionately with the circumstances and evidence.

However, distinctness can be problematic when a proposition conflicts with other beliefs.   Davidson (2004) gives the example of walking home to his house in Princeton and accidently walking into his neighbor's house. He describes the set of houses as being identical architecturally, although mirrored in their layout. Walking over the hearth, he perceives aspects of his neighbour's house that are incoherent with his memories of his house. The furniture is not right and he finds the neighbor's wife (rather than his own wife) in the kitchen and so forth. Initially, he rationalizes that his neighbor must be visiting him and that his own wife must have bought new furniture. So set in his belief that he has walked into his own home (how could he not!) that he invents a post hoc justification for each disconfirming piece of evidence that would ensure coherence amongst his beliefs.

A desire for coherence kept Davidson from the truth far longer than his pride and bashfulness would prefer. The fastest means for Davidson to get into the comfort of his own home would be to immediately doubt that he had arrived home upon seeing his neighbor in his kitchen. (After all, it would be a very rare occurrence indeed if she were there, alone in the afternoon.) Bayesian principles

offer the mechanism by which to have appropriate doubts: Very unlikely evidence given the hypothesis is a good justification for reconsidering the hypothesis.

This example shows the value of holding incoherent credences simultaneously. Davidson is rational to feel that he is home (after all, he just walked down his street). He is also rational to think that his neighbor's wife is in his kitchen (after all, there she is!). He is rational to suppose that he is in his neighbor's house (he can see his neighbor's furniture). He is also rational to imagine that his wife has bought new living room furniture (it is within the realm of possibility that she would do this).

It is rational for Davidson to hold many credences simultaneously that do not cohere as long as he attempts to resolve this incoherence for the sake of his goals. In this instance, his goal is to sit in his own armchair as quickly as possible. To reach that goal most effectively, he can apply Bayesian hypothesis testing with his evidence and derive the correct response—that he ought to apologize to his neighbor and walk directly out the door to his actual home next door. Choosing the right hypothesis depends on distinctness.

Incoherent sets of beliefs at different degrees are a normal state of play for any epistemic agent. What makes them defensible is that they are arrived at through reliable foundational means (e.g. perception, memory) or reliable intellectual competencies (e.g., Bayesian updating, causal reasoning). The rational being decides which of these incoherent sets to resolve in order to achieve her goals. Beliefs that are irrelevant to an agent's goals may remain distinct for the time being, but the rational being aims for epistemic homeostasis: the attainment

of true integrated beliefs optimized for their goals. Incoherence and distinctness are necessary and valuable parts of epistemic endeavor.

A critic of distinctness might say that distinctness is surely not a valuable epistemic good. Sometimes we consider a subset of beliefs regarding a particular problem, but this is a rhetorical method and not an evaluable end. Those methods are valuable only so far as they produce beliefs that cohere with our other beliefs. We do not value the blurry image produced by a short-sighted eye; no, we wait for the owner to put on glasses and give their pronouncement. We do not value the act of putting together a microscope (although we may respect it), but we value the completed object working well and calibrated correctly. The reason why epistemologists have not included distinctness as a valued epistemic good is that its goodness is entirely based on its role once no longer distinct. The parts of a microscope are valueless until we know they function cohesively together.

My response is: There no doubt traditional epistemology is focused on ends, not means, and in such a framework, distinctness is not particularly valuable. However, distinctness is highly relevant in a competence-based epistemology, such as a Bayesian virtue epistemology, which values *the way* agents come to hold beliefs and rationalizes *each* decision made. Bayesians value action under uncertainty: Has the agent chosen the most rational odds, given their circumstances? Virtue epistemology includes the processes by which agents come to hold beliefs as part of their definition of knowledge. If methods and outcomes are valuable, then distinctness is examinable and valuable.

3.8.1 FAILURES ARE A SOURCE OF KNOWLEDGE

> A single failure… is a source of knowledge we might not have gained in any other way… [They reveal] weakness in reasoning, knowledge and performance that all the successful designs may not even hint at… the best way of achieving lasting success is by more fully understanding failure. (Petroski, 1985)—Quoted in the New York Review of Books June 2013, LX(11), p.53.

Why are failures a source of knowledge? Why are skeptical attacks and counterexamples 'good practice' in epistemology? Failures are a source of knowledge because they a) allow us to identify the scope of what we did know, and b) help us adjust our parameters to know more in the future. Take the example of children constructing with sticks. They learn through experiment and failure that two sticks cannot stand on their own, but that three sticks mutually support one another in a teepee-type shape. Consider this a low level of knowledge. Through further experimenting, the children observe that the sticks cannot withstand downwards pressure from a fourth stick. They must reconsider their sticks and their aims before embarking on their next attempt. If they aim to build a fire, then the fourth stick is best placed within a teepee-like shape, which will strengthen it. However, if their aim is to build the tallest or strongest structures, their placement of subsequent sticks may diverge.

The point here is that as children progress and fail, they achieve certain aims and create new aims, and it is the same with knowledge: One starts with low levels of knowledge and build up to ever more comprehensive knowledge in certain domains. Knowledge is built by failing under skeptical pressure and through various internal and external permutations in conditions. There are limits to human knowledge, constrained by our physical selves, but each level of knowledge, rising from animal to reflective, is fully described within a particular

context, serving a particular knowledge need. Building knowledge requires perseverance, experimentation, failure, and feedback. With knowledge comes greater reliability and confidence, in a positive feedback cycle, which builds competence and capacity.

In the buzzing blooming confusion (James, 1890) of early life, agents are immersed in their environments physically. Mathematical properties of three-dimensional space are learnt geometrically by the splaying and retraction of limbs. Babies see unclearly, but they grasp and come to know embodied. The triangulation of multi-modal data and feedback through hypothesis testing is surely how knowledge is forged and first justified. Coordination between these sensory and propriosensory systems is vital for the success of the organism in their environment, but how does it work?

3.9 COORDINATION AS A SOURCE OF JUSTIFICATION

In this chapter, I have presented the theory that coherence is a source of justification when it emerges from satisficing or optimal relations between choristic sets of beliefs. I have also defended the justification of these sets as distinct from the rest of one's beliefs. At this point, one might reasonable ask what justifies the set up and dissolution of these sets? Is there a need for a third kind of justification: reliability, coherence, and coordination? Is a belief in any particular context justified not just because it is reliably formed and coherent, but also because of how it is coordinated between local and holistic goals?

The difficulty of coordination is well known in artificial intelligence through their attempts to get a legged robot to move over unpredictable surfaces

such as ice, mud, and rocks. To adapt and recover from obstacles and challenges, the 'Big Dog' robot responds quickly to adjudicate input from 50 sensors including: the motion of the body (acceleration and attitude) and joints; engine speed, temperature, and hydraulic pressure; the position and force the joints and the velocity and altitude during locomotion; and so on. (Raibert, Blankespoor, Nelson, Playter, & Boston Dynamics, 2013) Big Dog has robust fault tolerances due to the difficulties associated with managing the complexity of data. Coordination means that data from different inputs works together and supports other data to ensure performance success, and better coordination means faster progress across unpredictable surfaces. Importantly, the robot does not become more coordinated just through an increase of data. It is the way the data is managed and utilized that dictates its usefulness.

Coordination also means more than just mutual support—it means suppressing some data and communicating other data to ensure the stability of a system, and it results in convergence on the right action. The asymptotic convergence of error signals is crucial to ensure system stability and improve performance for mobile robots. So how does coordination relate to weak foundationalism and coherence?

Coherence for robots relates directly to performance over a given terrain or context. The degree to which they know the surface they are moving across depends on foundational data and the convergence of that data adjusting for errors. Sosa argues that epistemic virtues are dispositions to behave in ways that are likely to hit on the truth. The formation, maintenance and adjustment of these

dispositions stems from the convergence of error-prone systems and their divergence from data irrelevant to the task. Coordination justifies sets of beliefs, acknowledging that there is no way that *all* a person's beliefs could possibly reach consistency (Plantinga, 1993). As a result, any epistemology that depends on holistic logical consistency or advocates for it is untenable.

If information is not holistically integrated or encapsulated how might the mind interact with it? David Lewis (1968) set up the problem of coordination for cooperative decision-making. Take the dilemma of trying to call another person who is also trying to call you. Both parties cannot pick up the phone or leave the phone without failing to coordinate. As another example, consider some campers looking for firewood:

> It matters little to anyone in which direction he goes, but if any two go in the same direction they are likely to cover the same ground so that the one who gets there later finds no wood. Each must choose a direction to go according to his expectations about the others: one different from anyone else's (Lewis, 1968)

In this example, each person must modify their actions to promote shared interests. Lewis used his theory of convention to resolve social coordination problems, and although I do not draw on his social coordination solution (or justification) here, let us instead transpose the same issue of coordination to the mind. Suppose each of the mind's faculties behaved like self-interested entities with a shared agenda. A partial conflict of interest exists between these sub-systems; they have to compromise to produce optimal behavior—an *impure coordination problem* (Rescorla, 2011). Sets of beliefs coordinate like crew on an air raid. Pilots, gunners, bombers, photographers, and navigators have their own preferred trajectory with a shared desire to expel their arsenal and return safely to

base. They must coordinate with the rest of the crew to optimize the outcome for all.

Imagine the difficulties that beset the hunter on horseback during a storm. Her capacity to shoot arrows depends on output from the visual system, auditory system, and vestibular system amongst others. The visual system may return messages regarding prey ahead that conflicts with grunts heard off to the left. The muscles of the body rely on the horse's steady gait in order to line up a shot. The hunter is hungry and tired and wants to finish quickly, but she wants to shoot a male deer rather than a nursing mother. Each system is updating its beliefs continuously, and outputs may not cohere perfectly with one another to decide when to shoot one of a limited set of arrows. How might these systems coordinate when the aim is to accurately and discriminately shoot as quickly as possible?

Does the mind really have self-interested components? Can a coordination theory relevant to social convention (also decision theory) be ported across to justification? My approach to coordination is consistent with the Bayesian treatment of coherence in that: 1) Bayesian coherence is based on faculties of the mind operating like independent witnesses (Lewis, 1946); and 2) Bayesian coordination algorithms are already researched in neuroscience and implemented in robotics. I consider two coordination alternatives below—one from neuroscience and the other from engineering.

Recent work in dynamic coordination in the brain (von der Malsburg, Phillips, & Singer, 2010) offers a framework for coordination based on how the mind maintains flexibility and reliability:

> Neural activity…is dynamically coordinated so as to produce coherent patterns of macroscopic activity that are adapted to the current context, without corrupting the information that is transmitted by the local signals…
>
> …We need to understand better how these diverse activities [adaptation to different environments and tasks between brain regions, groups of cells and cells] are coordinated. Most locally specialized streams of activity can affect most others, either directly or indirectly. In this sense, the brain operates as an integrated whole. It is also clear that percepts, thoughts, and actions are normally, in some sense, both coherent and relevant to the current context.

The need to combine flexibility with reliability means that beliefs need to be used in different ways at different times and in different contexts. I do not have the space to go into dynamic coordination in great detail here, but I hope it is sufficient to suggest that there is a burgeoning field in studying coordination in the brain that might extend our understanding of the justification of beliefs.

Separate to biology, there has been work in engineering to enable limited separability of beliefs under the banner of enabling multiple autonomous entities to interact in a distributed, asynchronous manner (Pfeffer & Tai, 2012). This work concentrates on the practical problem of how to coordinate robots, but the slightly orthogonal domain should not put us off. In the same way that research in testimony helps us consider faculties of the mind, the outcomes for separate robots might be analogous to the fairly autonomous and asynchronous workings of human mental faculties. This work may further the goals of holism and informational integration (Weiskopf, 2008). It turns out that asynchronous updating works because the entire system relies on reflective beliefs formed earlier than current beliefs held at any one location. As Pfeffer & Tai (2012) state:

> Even with this scheme, we still cannot use the beliefs at the current subnode as the beliefs about the system state. Instead we have to use beliefs at a historical subnode. The problem is that the current subnodes will always have just received their first messages at the time they form their first beliefs. Thus they will not have had time to

converge to the correct beliefs. In contrast, historical subnodes will have received multiple sets of messages that will have traveled around the loops several times. The belief of a historical subnode about the state of the system at its time point will be much more accurate than the belief of the current node about the current state. …
There is a natural tradeoff here. Older reporting subnodes will have had more chance to converge to the approximately correct beliefs and will be more accurate about the state at their time point. On the other hand, as nodes get older, their beliefs become more stale and become a less accurate reflection of the state at the current time point In our experiments, we find that the second most current subnode is the best one to use.
One might think that using beliefs about a historical state as a substitute for beliefs about the current state is a drawback of our approach. However, this is unavoidable for any system that updates beliefs in real time. … It is inevitable that in any system that takes time to process information, the information will be old by the time it has been processed.

This account resolves the coordination problem by accepting the considered output of slightly older beliefs. Pfeffer & Tai's (2012) asynchronous algorithm has two advantages: 1) it avoids the need for a centralized controller (good for animal knowledge) and 2) it does not need synchronization (holistic coherence) in monitoring dynamic systems. Their algorithm assumes the entire system is changing even as updates occur. Each update requires small levels of computation and limited capacity to communicate at each node. Evidence is exchanged across nodes in the network imperfectly, in the sense that updating in one part of the system prevents another part from immediately benefiting from its processes. Nevertheless, eventually, evidence is propagated around the network.

Adding coordination as a source of justification is meant to shift the focus of debates in coherence beyond notions of 'mutual support' or 'consistency'. Somehow, agents come to know their world. They are justified because they get factual feedback with enough coherence such that they can hold beliefs. The success of these beliefs can be seen in the thriving of some organisms and the failures of others. Factors that affect an organism's capacity to manage data will

affect those organisms' capacities to have appropriately rational dispositions and this will be evidenced in their failure to achieve goals such as navigating over a particular terrain.

What emerges in this chapter is that normative coherence is perhaps best thought of as another form of reliabilism—a set of mechanisms or processes that govern belief-to-belief relations. Coherence includes the rules and processes that establish defensible relations between beliefs. This is a good outcome for Sosa's virtue epistemology, because Sosa wants to include coherence in his view, but notes that he had not decided precisely what role coherence might play.

"I leave open the question of whether the nature of coherence, and of understanding/explanation, requires explanation in terms of reliability in the actual world. Even if such explanation would be required at bottom, it may still be that coherence is a distinctive value with its own special status" (Sosa, 2009, V.2, II. Epistemic Values and Why Knowledge Is a Matter of Degree).

Foundationalist and coherentist justification are both ultimately based on reliabilist principles. That is, perception, memory, testimony, inference, and coordination need to be reliable in a particular context in order to justify belief. Virtue epistemology is right that the relevant measure of reliability is the competence of the agent's processes, rather than some externalist measure.

However, coherence is a set of processes that is distinct from foundational processes. Whereas foundational processes primarily concern the relationship between non-beliefs and beliefs, coherence relate beliefs to one another. In addition, coordination may be another layer of reliabilist justification, in that coordination processes manage the relevance and flexibility of foundational and coherence processes.

3.10 CONCLUSION

In this chapter, I have outlined a number of different theories of what coherence might be. Pure coherentism suffers from both the isolation and alternative systems objections. Weak foundationalism requires weak coherence— and even this is hard to defend. A Bayesian formulation of weak foundationalism produces dubious results including that highly cohering sets of propositions are less likely to be true than less coherent sets of propositions.

In response, I have argued that formal attempts (to date) to define coherence have not captured the sense in which coherence adds justification to reliably produced beliefs. Coherence is the inferential fastenings that hold beliefs together. These fastenings might be explicit beliefs, implicit beliefs, or unrepresented processes that weave beliefs together into functional sets in order for an agent to succeed in a particular context towards a particular goal. Coherence understood in this way raises the likelihood that a single belief (or hypothesis) is true by binding (entailment, inference, explanation, and so on) limited sets of propositions. Agents are reliable in so far as they coordinate the coherence and distinctness of their beliefs, and the skepticism and confidence with which they hold those beliefs. The mind will adopt principles of both coherence and distinctness as it moves through a variety of hypotheses. There is not an unlimited number of ways propositions can cohere under a weakly foundational account so described, because beliefs (or processes) based on causal inference will restrict the set of plausible alternatives.

When our performance is missing its mark, we need to reassess our patterns, and once a new strategy is confirmed, we can yet again set forth valuing coherence within a new paradigm. In this way, coherence and disintegration both act on the agent in turn, like the catabolic and anabolic forces within a cell. Ideally, the agent's epistemic structures operate like homeostasis: Self-regulating to maintain the optimum health of the organism.

4   HOMEOSTATIC EPISTEMOLOGY.

> In general, coordinating interactions are those that produce coherent and relevant overall patterns of activity, while preserving the essential individual identities and functions of the activities coordinated. (von der Malsburg, et al., 2010, p. 1)

4.1   INTRODUCTION

This dissertation is a study of virtue epistemology (Sosa, 2007, 2009a, 2011) and, in particular the competencies that produce rational credences and knowledge in the service of human flourishing despite the inevitable uncertainties an agent faces and the myriad limitations both on human cognition and the availability of information. Homeostatic epistemology supposes that an agent has a *rational credence p* when *p* is the product of reliable processes aligned with the norms of probability theory; whereas an agent *knows* that *p* when a rational credence *p* is the product of reliable processes such that:

1. *p* meets some relevant threshold for belief (such that the agent acts as though *p* were true and indeed *p* is true),

2. *p* coheres with a satisficing set of relevant beliefs and,

3. the relevant set of beliefs is coordinated appropriately to meet the integrated aims of the agent.

This chapter examines whether homeostatic epistemology can resolve some of the issues originally raised for virtue epistemology regarding the sufficiency of apt belief: In §4.2, I motivate the use of the term 'homeostatic epistemology' and clarify its commitments; in §4.3, I show how homeostatic justification might extend the usefulness and applicability of Sosa's 'apt belief' by defending the 'aptness' of apt belief, explaining levels of knowledge, limiting the risk of infinite

regress, and addressing the problem of luck and the value of testimony. I conclude the dissertation in §4.4.

4.2   HOMEOSTATIC EPISTEMOLOGY

Physiologists have long understood how important of the regulation of *milieu intérieur* (or the internal environment) is for the regulation of bodily functions (Bernard, 1858). The term 'homeostasis' was coined by an American physiologist Walter Cannon, to describe how the body maintains relative stability in bodily tissues that are critical to cell survival—e.g., nutrient availability, oxygen availability, temperature, pH, and ion concentrations (Cannon, 1929, 1939; Day, 2005)—and it refers specifically to the range of nested, self-regulating physiological processes that aim to achieve equilibrium within the body, despite varying internal and external conditions (Schulkin, 2003). Although homeostasis operates at the cell level, it involves many organs—"brain and nerves, the heart, lungs, kidneys and spleen, all working cooperatively" (Cannon, 1939)—and the processes are both local (e.g. heart, kidney) and broadly coordinated (e.g. by the brain) to maintain internal viability.

Additionally, homeostastic processes are intrinsically normative (Schulkin, 2003), aiming to improve the wellbeing of the agent, which links back to the Aristotelian notion of flourishing (that is, the best explanation of why the body's systems function the way they do is that they are calibrated to strive for optimal functionality given the constraints faced by the individual). The degree to which a physiological process is competent is, in essence, the degree to which it contributes to the body's optimal functioning. As an example, achieving a rapid

heart rate through athletic exertion is an appropriate way to both escape immediate threat and improve one's aerobic fitness, and as a result increase one's lifespan (Yataco Md, Fleisher Md, & Katzel Md, 1997). However, a rapid heart rate could also be a symptom of Graves' disease,[66] and in this particular circumstance it does not contribute to optimal functioning and instead leads to the loss of one's normal heart rhythm (atrial fibrillation), which in turn may lead to a stroke (Biondi & Kahaly, 2010). In summary, homeostatic processes are competent only when they operate reliably and coordinate across many of the body's systems to improve wellbeing.

Just as the heart reliably pumps blood to meet the needs of an animal during times of rest and times of intense activity, the mind reliably forms beliefs in different contexts[67]. Like homeostatic processes in the body, many of the mind's processes strive to put the agent in circumstances that are most conducive to their wellbeing.[68] The primary epistemic driver for an agent is achieving true beliefs about external and internal conditions that are relevant to their success. However, due to the limits of human cognition and the poverty of information in any given situation, there is also a secondary epistemic driver—achieving rational degrees of belief given one's evidence and circumstances (which includes withholding belief or maintaining low degrees of belief in a hypothesis). For example, even if

---

[66] Graves disease is a common form of hyperthyroidism in young women that affects homeostasis in a range of systems including circulation (heart palpitations), affect (nervousness), digestion (diarrhea, weight loss, increased appetite), skin (excessive sweating), and the muscular system (tremor and muscle weakness) (Hollander & Davies, 2009)

[67] An agent might reliably use their hearing to detect a fire alarm and reliably reason that the sound indicates a fire drill—because inductively, most times they have heard a fire alarm in that building, it has been an exercise, a prank, an accident or a fault, rather than evidence of an actual fire. The agent reliably reevaluates their evidence that the fire alarm signals a drill if new evidence relevant to their belief occurs, such as the smell of smoke (that has an prior probability close to zero if there is no fire). The agent does not alter their beliefs about the fire drill if they have evidence irrelevant to the existence of the fire (e.g. they smell perfume). In this way, for any context, there is a limited set of beliefs that are relevant to the justification of a single belief and can elevate that belief to knowledge.

[68] Sosa (2011, p. 57) argues that beliefs aim for truth. I argue that it is not the beliefs per se, but the homeostatic processes that control the production, coherence and coordination of beliefs that are truth-seeking.

an agent rationally believes that there is no fire despite being able to hear a fire alarm sounding, they maintain an appropriate caution in case their inductive inference is not justified in this particular instance.

Epistemologists have long wondered how the mind regulates beliefs to achieve correspondence with the world, and what justifies these processes. Two prevalent conceptions of justification include reliability (that belief-making processes consistently make true beliefs) and coherence (that beliefs gain justification to the degree that they support other beliefs).[69] I take Quine (1970) to be correct when he asserts that any belief might be affected by any new belief, but that justification must be limited to *choristic* sets (partitions of cohering beliefs that add justification to foundationally justified beliefs) where no further explanation is needed (see §1.2, §3.1, §3.5, and §3.6). The homeostatic *choristic* solution reconciles the following statements: 1) knowledge is common and ubiquitous, 2) people hold inconsistent and false beliefs, 3) people hold beliefs that do not abide by normative coherentist (inference) rules, and 4) all beliefs are part of a web of belief. Justification by limited sets explains how agents know about the world, but also how they cope with uncertainty, incomplete evidence, inconsistent or false beliefs, and changing environments.

---

[69] Other sources of justification include a priori and evidential. A priori beliefs are justified in so far as they arise from processes independent of experience "*beyond that needed to acquire the concepts required to understand the proposition at issue*" (Russell, 2013)(author's italics). My arguments do not depend on disproving a priori theories of justification, though I err on the side of doubt that a priori justification is genuinely explanatory. Evidential justification supposes that the presence of perceptual, introspective, memorial, and/or intuitional experiences justify beliefs (e.g. the fact a cup of coffee tastes sweet is evidence that it is sweet). Reliabilists agree that evidence plays a role, but only if the experience arises as a result of reliable processes, rather than the experience per se (e.g., a bent-looking stick is not evidence that it is bent if the stick is partially underwater and viewed from above).

Homeostatic knowledge is achieved when localized sets of beliefs function within narrow constraints and coordinate across different contexts and circumstances to converge on the right action. These beliefs may be quasi-apt beliefs (rational credences) or apt beliefs that confer knowledge within a specified domain. Like the organs and biological systems of the body, rational credences and apt beliefs operate in functional sets that may or may not relate to one another, depending on the circumstances. Epistemic homeostasis occurs when an organism coordinates sets of beliefs to achieve their aims. The project of homeostatic epistemology can shed valuable light on epistemological issues raised earlier in this thesis.

4.3   HOMEOSTATIC JUSTIFICATION

4.3.1   APT BELIEF IS APT ENOUGH

In §2.3.1, I considered the argument that apt belief is not 'apt enough' to count as knowledge, ordinarily conceived (Lepock, 2010) which Lepock argued is robust against a myriad of skeptical concerns that simple reliabilist animal beliefs fail to adequately respond to. I suggested, however, that apt belief is sensitive and flexible enough to adapt to difference circumstances, which makes it plausible that apt belief is sufficient for ordinary knowledge. I then argued, in §3.9, that reliabilist processes include not only *foundational* processes, but also the subtle and complex processes of *coherence* that bind together foundationally acquired beliefs, as well as the integrative and dissociative processes of *coordination* that determine what sets of beliefs are relevant to the agent in the circumstances they are in. The creative, flexible, adaptive quality of apt belief may stem from

processes of dynamic coordination between sets of foundational, coherent beliefs. To be efficient and dependable, beliefs must arise predictably when used at different times and in different contexts; but to be flexible and creative, beliefs must arise idiosyncratically at different times and in different contexts. Dynamic coordination is the study of the processes that combine and sequence beliefs in a context-sensitive way, changing from moment to moment, even though the beliefs that are coordinated retain individual or group justification (von der Malsburg, et al., 2010, p. 2). A good introduction to dynamic coordination is the effect of attention on beliefs.

Consider the difference between seeing a hiking trail ahead of you and focusing on a particular part of the trail, say a footprint on the track. It is not difficult to imagine the typical absent-mindedness that occurs when one is hiking, which means that although we see the path ahead, we see it in a trivial way. The act of paying attention to something in your visual field does not radically change your visual percept, i.e. the same sorts of colors, shapes, and textures are processed. But, paying attention can have a large impact on how you think about what you see. Suppose the footprint that you saw entered your foveal area just before you heard an unnerving guttural sound. Once paying attention you would analyse the paw print carefully—is it a brown bear or a black bear? Are these tracks fresh? The act of paying attention to a visual stimulus is an example of dynamic coordination because it alters the way perceptual information is processed in the visual cortex—e.g., increasing sensitivity to faint stimuli, or

reducing the impact of task irrelevant distractors, while preserving the original inputs (Reynolds & Heeger, 2009).

Attention is particularly relevant to a discussion of apt belief because—as discussed in §1.6—attention can be a reaction to stimuli (e.g., orienting to a threatening sound) or a conscious choice (e.g., focusing on footprints on the path). The former is typical of the animal knowledge produced by System 1, whereas deliberative attention is typical of reflective knowledge and is produced by System 2 (Kahneman, 2011). Although dynamic coordination occurs for both animal and reflective knowledge, reflective knowledge engages a level of analysis that is more comprehensive than the coordination of animal knowledge. These findings are good news for Sosa's virtue epistemology. Interpreted homeostatically, the 'adroitness' in Sosa's AAA knowledge structure consists of foundational, cohering, and coordinating competencies that ebb and flow depending on the agent's circumstances, needs, and limitations.

4.3.2 LEVELS OF KNOWLEDGE

In §2.3.2 I used Sosa's trickster (2007) to demonstrate that the level of our knowledge at any particular moment varies depending on our goals and capacities. For example, a child, an art critic and a magician's assistant have different epistemic priorities when they view a table in unusual lighting conditions, and their success ought to be measured relative to their interests. The set of beliefs required for a child's animal knowledge that the table is red is different from the set of beliefs required for the magician's assistant to reflectively know that the table is red. However, both the child and the magician's

assistant will be similarly justified in believing that the table is red at the level of animal knowledge. It is also possible that an individual's epistemic priorities may change over time. For example, a father and son at an art gallery might view the red table and share animal knowledge of its redness. The son may then ask his father why the table looks red, prompting the father to reflect on the question and direct his attention more closely in order to understand how the table has been made and the origin of the redness. Based on his inspection, the father might deduce that the table looks red because there is red LED lighting within what is actually a clear laminate table.

I believe that at any one point in time, an agent has animal knowledge about a variety of different things in their internal and external environment. Reflective knowledge is attained when beliefs that constitute animal knowledge are integrated into larger, choristic sets that include beliefs about why the animal knowledge is truth-conducive. In the case of the father, he has animal knowledge that the table is red and reflective knowledge that the human visual system will produce the experience of redness when red LED lights are switched on inside the table and external lighting is normal white light. He might also hypothesize with his son that the table could still look red if the internal LEDs were changed to white and the external lighting was changed to red. Theoretically, reflective knowledge could become iteratively more comprehensive if the cohering sets of beliefs that are relevant to an agent's goal incorporated even more explanation. However, as an agent's explanatory goals become more broad-ranging, and the size of the choristic set becomes larger, the risk of falsehoods within the set

increases—as warned by Klein & Warfield (1994) in §3.5—which places the coherence of the entire set at risk and therefore threatens its justification.

4.3.3 REFLECTIVE KNOWLEDGE

In §2.4, I began my analysis of reflective knowledge by considering Sosa's application of the *principle of exclusion* PE:

> PE: if one is to know that h, then one must exclude (rule out) every possibility that one knows to be incompatible with one's knowing that *h*" (Sosa, 1997, p. 411).

However, PE seemed too permissive of merely conceivable possibilities, rather than likely alternatives, and for this reason, I argued that reflective knowledge must only be defended against incompatible possibilities that cohere with the rest of one's beliefs by the *principle of exclusion and coherence* PEC:

> PEC: if one is to know that h, then one must exclude (rule out) every possibility that one knows to be incompatible with one's knowing that h *that coheres with the rest of one's beliefs*.

I then reconsidered reflective knowledge from a Bayesian perspective (in §2.5) and transformed PEC into the *principle of exclusion and coherence from multiple hypotheses* PECMH:

> PECMH: if one is to know that h, then h must be the most likely amongst a set of plausible hypotheses h_m to h_n that cohere with the rest of one's beliefs at lower likelihoods.

After analyzing the role of coherence in §3.5, §3.6, and §3.7, I redefined reflective knowledge again based on a *principle of exclusion and _relevant_ coherence from multiple hypotheses* (PERCMH):

> PERCMH: if one is to know that h, then h must be the most likely amongst a set of plausible hypotheses h_m to h_n that cohere with *the relevant partition of one's beliefs* at lower likelihoods.

While both animal and reflective knowledge involve limited, mutually supporting partitions of beliefs, reflective knowledge includes higher order beliefs—including beliefs about abstract concepts, causality and theoretical entities or relations as well as particulars within a domain that explain animal knowledge (Gopnik & Wellman, 2012). Hierarchically nested probabilistic methods and other inferential fastenings weave beliefs together into functional sets in order for an agent to attain reflective knowledge in a particular context, towards a particular goal.

When Sosa says that he wants reflective knowledge to encompass the capacities of the intellect, information and deliberation, he calls for an account that acknowledges that the mind can hold multiple hypotheses simultaneously, and adjust credences, as needed, in response to deliberation and information. The mind adopts principles of both coherence and distinctness as it considers a variety of hypotheses according to the norms of Bayesian reasoning (see §2.6.4) which includes: anchoring your judgment in base rates relevant for the decision, questioning the diagnosticity of your evidence, and remaining open-minded, adaptive, and objective. Homeostatic epistemology can strengthen Sosa's reflective knowledge against the wrath of rationalists seeking to vindicate the intellect.

One of the advantages of limiting coherence to a partition of beliefs is that counterfactual reasoning and conjecture can take place distinct from other beliefs. In this dissertation, I have made the case that distinction should be viewed as a valuable part of epistemic justification (§3.8). If beliefs are coordinated

dynamically, then justified partitions of beliefs can easily be taken offline, decoupled for intellectual hypothesis testing, or recoupled to sets of beliefs for empirical testing in the world—as occurs during exploratory, pretend, and imaginary play (Gopnik & Wellman, 2012). Such an account may conflict with current theories of pretend or imaginary play that involve quarantining a silo of pretend beliefs from normal beliefs (Leslie, 1987; Nichols & Stich, 2003) and this would be an interesting avenue of future (and possibly controversial) research.

4.3.4 THE RISK OF INFINITE REGRESS

In §2.4.1 I considered the Pyrrhonian argument that an infinite number of beliefs are needed to justify any belief. I argued that the regress of aptness and meta-aptness was not a threat to virtue epistemology, but rather an explanatory tool to describe different levels of competence and performance domains. In §3.5, §3.6, and §3.7, I elaborated on this argument by claiming that, in fact, a maximal fittedness is required for any set of beliefs for any particular goal of the agent in a particular situation[70]. Once a set of beliefs has met a minimal satisficing level of cohesion, a belief within the group would be sufficiently justified to count as knowledge for that agent, in that context, and the degree of justification required for knowledge ought to be limited by the subject's computational, attentional, and inferential limits.[71] Knowledge is objectively evaluable, but the scope of knowledge must be established relative to the capabilities of the agent (or perhaps normal human capabilities). Consider what a butterfly knows as it approaches a flower. Butterflies and humans can detect red, blue, and green, but butterflies can

---

[70] While good in theory, I acknowledge that we still do not understand the mechanism by which the relevant set of beliefs is decided (see §1.73), only that some subset must suffice given the presence of knowledge in our day-to-day experience.
[71] Thanks to Ernest Sosa for highlighting this condition on justification (personal communication, September 3, 2013).

also see the ultraviolet spectrum. A human can know that the flower *helianthus tuberosus* is yellow, even though a butterfly would have a more nuanced color experience that included yellow and purple.

4.3.5 LUCK

In §2.4.2, I considered the case of double luck—i.e., Gettier style cases—and I described a scenario in which a highly skilled archer's arrow lands on the target due to a double gust of wind. The intuition is that no one would accord the archer credit for this shot, even though it was accurate and adroit. Sosa explains the archer's failure to know by requiring aptness to be accurate due to the manifestation of competence—something a double luck case fails to exhibit—and I accepted Sosa's requirement, though I questioned the analogy. Beliefs are part of integrated sets of beliefs that mitigate against the type of luck described during the archer's performance. As such, I compared beliefs to a tomahawk missile with multiple levels of course correction and feedback. If a tomahawk fails in an environment it is designed to succeed in, then it is a failure of competence (a broken mechanism or maintenance error), rather than luck. The power of competencies to explain instances of knowledge is improved when they are in operation throughout the long sustained period of belief.

A critic might object to my counterexample, arguing that competently formed beliefs do suffer the ravages of luck, without a feedback mechanism, due to the temporal distance between forming a belief and executing a belief. Consider the following scenario:

> …not every case of knowledgeable belief requires a feedback mechanism, not even when the belief is a prediction, so that there is a time lapse between the formation of the belief and (in a relevant way) its hitting the mark of truth. You solemnly promise to meet me, and I believe you with high justification. And you do eventually show up at the appointed time and place. However, you had completely forgotten, and eventually showed up, not in order to meet me, but in order to buy some ice cream. Here my belief succeeds. It is correct. And it is competent. But it is not apt. Nor, in the circumstances, was there any feedback mechanism that could have corrected along the way.[72]

For the sake of clarity, let us call the *promiser* Pinocchio, the *promisee* Geppetto and the place Ponte Vecchio—the oldest bridge in Florence. In this scenario, both Pinocchio and Geppetto's beliefs must be justified by a set of beliefs that are relevant to the individual, and how those beliefs are coordinated over time—because a competent belief is never adrift from relevant cohering beliefs. Geppetto's belief that Pinocchio will arrive at the appointed time and place is based on other beliefs about Pinocchio's trustworthiness in this sort of situation, as well as beliefs about anything idiosyncratic that might make this situation different from previous experience. Pinocchio's solemn promise to meet Geppetto is justified if he truly intends to fulfill his obligation, desires not to lie, has the volitional capacity to follow through on his promise and so forth. However, what justifies Gepetto and Pinocchio's trust—long after the belief is formed—is also how the original belief is coordinated and connected with other beliefs and priorities *over time*.

Pinocchio and Gepetto's beliefs (formed at time t_1) are justified by, amongst other things, the relevant belief that Pinocchio's memory processes are sufficiently reliable to carry the belief 'across the line' to Ponte Vecchio at t_2. Reflective knowledge regarding the reliability of Pinocchio's memory processes

---

[72] Thanks to Ernest Sosa for his articulation of this scenario (personal communication, 12th August, 2013).

is a central presupposition of the solemn promise to meet at time t_2 and the justification of Gepetto's trust in the promise at t_1. Such a promise suggests that Pinocchio will prioritize this meeting over and above other considerations, and that the meeting constitutes a goal that trumps other competing agendas. After all, once memory processes fail and Pinocchio only accidently lives up to his prior commitment, his original belief and testimony are no longer justified—as evidenced by a loss of trust between Pinocchio and Gepetto.

Gepetto is only lucky if the memory error is unpredictable—for example, Pinocchio may have suffered a stroke that gave him amnesia. In such a case, Gepetto was justified holding the belief and the belief succeeds, yet he does not have knowledge, and this is why Sosa is right that there has to be the right connection between accuracy and adroitness. Gepetto must meet Pinocchio *because* of his competent assessment that Pinocchio would meet him, not because of luck. If Pinocchio's unpredictable stroke and hunger for ice cream was responsible for the meeting, it would not affect the trust relationship between himself and Gepetto[73].

However, what if the memory error is predictable? The purported equivalence of luck in this example is that Pinocchio had "completely forgotten" that he and Gepetto were supposed to meet, but forgetting is mostly a consequence of poor planning, rather than luck. Memory processes are a core part of the coordinating processes Pinocchio needs in order to arrive at the appointed time and place. If a craving for ice cream moves his muscles to Ponte Vecchio,

---

[73] Of course, the impact of the stroke may have additional complications for Pinocchio and Gepetto's relationship as they both must navigate Pinocchio's unfamiliar post-stroke cognitive capacities. Still, if both parties acknowledge their ignorance about Pinocchio's reliability, then they can maintain a trusting relationship.

the causal connection (and the competence of the original belief) is lost. I refer here to Martin and Deutscher's (1966) seminal work on the philosophy of memory to highlight the importance of the causal connection between a belief and its subsequent states. To *remember*, a person's original belief must be "operative in producing a state or successive states in him finally operative in producing his [memory] representation" (p.173). In this example, the failure to remember indicates that there is no longer a causal connection between arriving at Ponte Vecchio and the now vacuous promise to meet up.

Of course, Sosa (1997) is well aware that knowledge requires the right causal connection between the formation of a belief and the situation where that belief is operative. If memory systems fail to retain and utilize a belief established reliably and coherently, then it is a failure of knowledge due to a failure of coordination. If this analysis is right, then Pinocchio's original solemn promise to meet Gepetto is not fully justified for either Pinocchio or Gepetto because justification requires: a) that the belief was reliably formed, b) appropriate coherence within a relevant subset of beliefs, and c) that this belief is coordinated with the other cognitive, motor, and affective systems that enable the agent to get to the location because of their desire to meet. Pinocchio and Gepetto ought to have known that that Pinocchio is unreliable, even when he solemnly promises to meet up. This failure to know means that their belief that Pinocchio would meet Gepetto at the Ponte Vecchio was not adequately justified. Coherence and coordination further explicate the importance of a causal connection between adroitness and accuracy in Sosa's apt belief.

4.3.6 THE VALUE OF TESTIMONY

A key difference between testimony in virtue epistemology and Bayesian epistemology is the reliance on authorities and experts. Due to the high value placed on the transmission of knowledge—rather than credences—virtue epistemology (and epistemology traditionally) do not tolerate dubious testimony. Sosa (2011, p. 72) argues that the transmission of relevant information is the most salient property that explains the success of expert testimony (p.128). As such, the glory for success goes to the provider of testimony, rather than the receiver, whose only virtue is that she trusted the provider. For testimony to count as knowledge, it must be accurate, and it must be accurate in virtue of expert competent processes. However, the most salient reason for the testimony being right may not be the fact that it arose from the expert. A belief's correctness may stem instead from the cognitive accomplishments of others. This point is recognized by Bayesian approaches to testimony, which include a comprehensive calculus for weighing testimony according to various criteria, including the prior probability of each hypothesis and the varying reliability of witnesses. A useful epistemology will certainly value highly reliable, expert testimony more than less reliable testimony, but testimony given by unreliable witnesses can also be valuable, as long as there are more than one and their accounts are independently given (Bovens & Hartmann, 2004).

One of the problems with demanding that testimony is knowledge is that both the believer and the expert may not have any capacity to determine whether the testimony is knowledge at the time it is trusted. The expert has a high degree of confidence in their belief, and the believer accepts this confidence when the

belief is transmitted. If the belief happens to be knowledge, then it benefits the believer; if the belief turns out to be a justified but untrue belief, then it is an unfortunate accident, rather than an indication that the expert ought not be trusted on some future occasion. The point is that the Bayesian contribution offers virtue epistemology an excellent way to manage the fallibility of reliable processes that still improves truth in the long term.

An additional benefit of a Bayesian virtue epistemology is that it recognizes the value of multiple witnesses. In a traditional epistemology, testimony does not, in itself, transmit justification. This is because beliefs conveyed through testimony have warrant in virtue of how they arose in an individual (e.g. perception, memory and/or inference). In a Bayesian virtue epistemology, however, cohering testimony from multiple witnesses *does* increase the justification of a belief. Justification arises from the intersection of agreement between at least partially reliable witnesses. I have not yet examined precisely how homeostatic epistemology ought to consider testimony, but this is another area that has great scope for future research.

4.4   CONCLUSION

I started this project as an empirical and philosophical examination of Ernie Sosa's virtue epistemology, particularly his division of knowledge into animal and reflective. I thought to myself: Is this right? Is this really what is really going on? Certainly, there were some good precedents for such a division in the memory and decision-making literature, particularly that learning mechanisms and decision-making can be split into implicit and explicit processes that drive belief

acquisition and updating. If animal knowledge accords with implicit memory and decision-making processes and reflective knowledge accords with explicit memory and decision-making processes, then it has empirical 'legs' (as it were). So far, so good.

However, Sosa claims, rather radically, that the same sorts of reliabilist justification that are relevant to animal knowledge also apply to reflective knowledge—an iterative reliabilism. I was intrigued by this idea of iterative reliabilism, but, like other philosophers, I was somewhat perplexed by how the high standard of reflective knowledge—incorporating traditional ideas about the intellect and higher order thinking—could be attained by reliabilist justification. It was not clear to me how one could attain reflective knowledge from simple reliabilist processes. In addition, reliabilist processes need an additional boost from coherence—an intuitively attractive, yet formally tricky, means of improving the veracity of belief. At the end of the journey, I feel that I have, at least to some degree, resolved this puzzle.

The justification of 'apt belief' and 'aptly apt' belief ought to incorporate apparatus from Bayesian decision theory. Bayesian processes operate for both implicit and explicit justification and they are grounded in reliabilist principles (i.e. they yield accurate results in the long term). Bayesian epistemology—as an agent-centered, competence-based, reliabilist epistemology—really gives virtue epistemology a strong framework to defend higher order reliabilist thinking. Normative coherence (coherence as a source of justification) is best understood as a variety of reliabilism. Importantly, our beliefs do not become more coherent

simply because we have more of them. Instead there is an ideal level of coherence that must exist between any number of propositions for some goal in a particular context. Coherence must be balanced with disintegration. Coherence and disintegration both act on the agent in turn, like the catabolic and anabolic forces within a cell.

Ideally, then, the agent's epistemic structures operate like homeostasis—self-regulating to maintain optimum health of the organism. Systems in dynamic equilibrium are under continual feedback control process to keep conditions relatively uniform. Reliability requires both skepticism and our trust in our faculties. We build up our capacities through our successes and failures as we interact with the world and it is epistemic homeostasis that ensures that the mind's belief-updating systems maintain equilibrium no matter the internal or external information environments. If homeostasis is the "wisdom of the body" (Cannon, 1939; Starling, 1923), then epistemic homeostasis is the wisdom of the mind.

5   REFERENCES:

Alxatib, S., & Pelletier, F. J. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind & Language, 26*(3), 287-326.

Armstrong, D. (1968). *A materialist theory of the mind*. New York: Routledge & Kegan Paul.

Ayer, A. J. (1956). *The Problem of Knowledge*. Bungay, Suffolk: Penguin Books.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science, 3*(1), 2-13.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances, from http://www.stat.ucla.edu/history/essay.pdf

Bealer, G. (1987). The philosophical limits of scientific essentialism. *Philosophical Perspectives, 1*, 289-365.

Bealer, G. (1993). The incoherence of empiricism. In S. Wagner & R. Warner (Eds.), *Naturalism: A Critical Appraisal* (pp. 163-196): Notre Dame University Press.

Bealer, G. (1998). Intuition and the Autonomy of Philosophy. In M. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 201-239): Rowman & Littlefield.

Bernard, C. (1858). Leçons sur la physiologie et la pathologie du système nerveux.   Retrieved from http://vlp.mpiwg-berlin.mpg.de/library/data/lit28997/index_html

Bes, B., Sloman, S., Lucas, C. G., & Raufaste, É. (2012). Non-Bayesian Inference: Causal Structure Trumps Correlation. *Cognitive Science, 36*(7), 1178-1203. doi: 10.1111/j.1551-6709.2012.01262.x

Biondi, B., & Kahaly, G. J. (2010). Cardiovascular involvement in patients with different causes of hyperthyroidism. [Report]. *Nature Reviews Endocrinology, 6*, 431-443. doi: 10.1038/nrendo.2010.105

Bishop, M. (2010). Why the generality problem is everybody's problem. *Philosophical Studies, 151*(2), 285-298. doi: 10.1007/s11098-009-9445-z

Blakemore, C., & Cooper, G. F. (1970). Development of the Brain depends on the Visual Environment. *Nature, 228*(5270), 477-478. doi: 10.1038/228477a0

Blumson, B. (2012). Mental maps. *Philosophy and Phenomenological Research, 85*(2), 413-434. doi: 10.1111/j.1933-1592.2011.00499.x

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

BonJour, L. (1999). The dialectic of foundationalism and coherentism. In E. Sosa & J. Greco (Eds.), *The blackwell guide to epistemology*. Malden, MA: Blackwell.

BonJour, L. (2003). Reply to Sosa. In L. BonJour & E. Sosa (Eds.), *Epistemic Justification*. Malden, MA: Blackwell.

Bovens, L., & Hartmann, S. (2004). *Bayesian epistemology*. Oxford: Oxford University Press.

Bower, G. H. (2000). A brief history of memory research. In E. Tulving & F. I. M. Craik (Eds.), *Oxford Handbook of Memory* (pp. 3-32). New York: Oxford University Press.

Bowers, J. S., & Davis, C. J. (2012). Bayesian Just-So Stories in Psychology and Neuroscience. *Psychological Bulletin, 138*(3), 389-414. doi: 10.1037/a0026450

Bradden-Mitchell, D., & Jackson, F. (1996). *The philosophy of mind and cognition*. Oxford: Oxford University Press.

Brainerd, C. J., & Reyna, V. F. (2002). Recollection rejection: How children edit their false memories. *Developmental Psychology, 38*, 156-172. doi: 10.1037/0012-1649.38.1.156

Buchsbaum, D., Bridgers, S., & Weisberg, D. S. (2012). The power of possibility: causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B., 367*(1599), 2202-2212.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews, 9*(3), 399-431.

Cannon, W. B. (1939). *The wisdom of the Body* (Revised and enlarged edition ed.). New York: W. W. Norton.

Carey, S. (2009). *The Origin of Concepts* Oxford, UK: Oxford University Press.

Carrier, M. (1991). What is wrong with the miracle argument? *Studies in the History and Philosophy of Science, 22*(1), 23-36.

Carroll, L. (1895). What the tortoise said to Achilles. *Mind, 4*(14), 278-280.

Chalmers, A. F. (1989). *What is this thing called science?* (2nd ed.). Brisbane, Australia: University of Queensland Press.

Cherniak, C. (1981). Minimal rationality. *Mind, 90*(358), 161-183.

Christensen, D. (2010). Rational Reflection. *Philosophical Perspectives, 24*(1), 121-140. doi: 10.1111/j.1520-8583.2010.00187.x

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204. doi: doi:10.1017/S0140525X12000477

Cohen, S. (1984). Justification and truth. *Philosophical Studies, 46*, 279-295.

Comesaña, J. (2002). The Diagonal and the Demon. *Philosophical Studies, 110*(3), 249-266. doi: 10.1023/a:1020656411534

Conee, E., & Feldman, R. (1998). The Generality Problem for Reliabilism. *Philosophical Studies, 89*(1), 1-29. doi: 10.1023/a:1004243308503

Conee, E., & Feldman, R. (2004). *Evidentialism*. Oxford: Oxford University Press.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin, 136*(1), 87-102.

Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.

Davidson, D. (2004). 12. Incoherence and Irrationality. Problems of Rationality. Oxford: Oxford Scholarship Online. Retrieved from http://www.oxfordscholarship.com/view/10.1093/0198237545.001.0001/acprof-9780198237549?rskey=v3dRlK&result=1&q=problems of rationality. doi: 10.1093/0198237545.003.0012

Day, T. A. (2005). Defining stress as a prelude to mapping its neurocircuitry: No help from allostasis. *Progress in Neuropsychopharmacology & Biological Psychiatry, 29*(8), 1195-1200. doi: 10.1016/j.pnpbp.2005.08.005

de Finetti, B. (1980). La Prevision: Ses Lois Logiques, Se Sources Subjectives (Annales de l'Institut Henri Poincare 7 (1937), pp.1-68 (H. E. Kyburg, Trans.). In H. E. Kyburg & H. E. Smokler (Eds.), *Studies in subjective probability*. New York: Krieger Pub Co.

Denton, D. A., McKinley, M. J., & Wessinger, R. S. (1996). Hypothalamic integration of body fluid regulation. *Proc. Natl. Acad. Sci USA, 93*, 7397-7404.

Devitt, M. (1997). *Realism and Truth* (2nd ed.). Princeton, New Jersey: Princeton University Press.

Devitt, M. (2002). Underdetermination and Realism. *Noûs, 36*(s1), 26-50. doi: 10.1111/1468-0068.36.s1.1

Devitt, M. (2006). *Ignorance of Language*. New York: Oxford University Press.

Devitt, M. (2011). Methodology and the nature of knowing how. *The Journal of Philosophy, 108*(4), 205-218.

Döring, F. (1999). Why bayesian psychology is incomplete. *Philosophy of Science, 66*(Supplement. Proceedings of the 1998 biennial meetings of the philosophy of science association. Part I: Contributed papers), S379-S389.

Dunn, J. S. (2010). *Bayesian epistemology and having evidence*. PhD, University of Massachusetts, Amherst. Retrieved from http://scholarworks.umass.edu/open_access_dissertations/273/

Eichenbaum, H. (2008). 3.01 - Introduction and Overview. In H. B. Editor-in-Chief: John (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 1-8). Oxford: Academic Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1997). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: The MIT press.

Evans, J., & Frankish, K. (Eds.). (2009). *In two minds: Dual processes and beyond*. New York: Oxford University Press.

Ewing, A. C. (1934). *Idealism: A Critical Survey*. London: Methuen.

Fantl, J. (2012). Knowledge how. *The Stanford Encyclopedia of Philosophy* Winter 2012. Retrieved 17 July, 2013, from http://plato.stanford.edu/entries/knowledge-how/

Feldman, R. (1985). Reliability and justification. *The Monist, 68*, 159-174.

Feldman, R. (2012). Naturalized epistemology. *The Stanford Encyclopedia of Philosophy* Summer. 2012, from http://plato.stanford.edu/archives/sum2012/entries/epistemology-naturalized/

Fivush, R., & Haden, C. A. (Eds.). (2003). *Autobiographical Memory and the Construction of a Narrative Self: Developmental and Cultural Perspectives*. Mahwah: Lawrence Erlbaum.

Fodor, J. A. (1968a). The appeal to tacit knowledge in psychological explanation. *The Journal of Philosophy, 65*(20), 627-640.

Fodor, J. A. (1968b). *Psychological explanation*. New York: Random House.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1983). *Modularity of mind: an essay on faculty psychology*. Cambridge, MA: The MIT Press.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. New York: Oxford University Press.

Fodor, J. A., & Lepore, E. (1996). The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition, 58*(2), 253-270.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

Forester, C. S. (1950). *Mr. Midshipman Hornblower*: Amazon Kindle Edition.

Friedman, J. (2011). Suspended judgment. *Philosophical Studies, 162*(2), 165-181. doi: 10.1007/s11098-011-9753-y

Furness, J. B., Clerc, N., & Kunze, W. A. A. (2000). Memory in the enteric nervous system. *Gut, 47*(Supplement 4), iv60-62. doi: 10.1136/gut.47.suppl_4.iv60

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, M.A.: The MIT Press.

Gallistel, C. R. (2012). Machinery of cognition. In P. Hammerstein & J. R. Stevens (Eds.), *Evolution and the Mechanisms of Decision Making (Strüngmann Forum Reports)*. Cambridge, MA: MIT Press.

Gallistel, C. R., & King, A. P. (2009). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. Hoboken NJ: Wiley-Blackwell.

Ganzel, B. L., Morris, P. A., & Wethington, E. (2010). Allostasis and the human brain: Integrating models of stress from the social and life sciences. *Psychological Review, 117*(1), 134.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gerrans, P. (1999). Delusional misidentification as subpersonal disintegration. *The Monist, 82*(4), 590.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2*, 83-115. doi: 10.1080/14792779143000033

Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. New York: Oxford University Press.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*(1), 107-143.

Gigerenzer, G., & Hoffrage, U. (2007). The role of representation in Bayesian reasoning: Correcting common misconceptions. *Behavioral and Brain Sciences, 30*(3), 264-267. doi: 10.1017/s0140525x07001756

Gigerenzer, G., & Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese, 187*(1), 243-268. doi: 10.1007/s11229-011-0030-6

Godfrey-Smith, P. (2001). Three kinds of adaptationism. In S. H. Orzack & E. Sober (Eds.), *Adaptionism and Optimality* (pp. 335-357). Cambridge, UK: Cambridge University Press.

Goldman, A. I. (1978). Epistemics: The regulative theory of cognition. *The Journal of Philosophy, 75*(10), 509-523.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. I. (1994). Naturalistic epistemology and reliabilism. In P. French, T. Uehling & T. Wettstein (Eds.), *Midwest Studies in Philosophy XIX: Philosophical Naturalism*. Minneapolis: University of Minnesota Press.

Goldman, A. I. (2008). Reliabilism. *Stanford Encyclopedia of Philosophy* Retrieved 3 October, 2012, from http://stanford.library.usyd.edu.au/archives/fall2009/entries/reliabilism/

Goodman, N. (1983). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Gopnik, A., Meltzoff, A. N., & Bryant, P. (1997). *Words, Thoughts, and Theories*. Cambridge, MA: The MIT Press.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085-1108. doi: 10.1037/a0028044

Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind, 115*(459), 607-632. doi: 10.1093/mind/fzl607

Greco, J. (2000). *The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry*. Cambridge, UK: Cambridge University Press.

Greco, J. (2010). *Achieving Knowledge, *. Cambridge: Cambridge University Press.

Greenberg, D. L., & Rubin, D. C. (2003). The Neuropsychology of Autobiographical Memory. *Cortex, 39*(4), 687-728. doi: 10.1016/s0010-9452(08)70860-8

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Gwin, M. F. (2011). *The virtues of Bayesian epistemology*. PhD, University of Oklahoma, ProQuest, UMI Dissertations Publishing. (3488133)

Hagura, N., Kanai, R., Orgs, G., & Haggard, P. (2012). Ready steady slow: action preparation slows the subjective passage of time. *Proceedings of the Royal Society B: Biological Sciences, 279*(1746), 4399-4406. doi: 10.1098/rspb.2012.1339

Hajek, A., & Hartmann, S. (2009). Bayesian epistemology. In J. Dancy, E. Sosa & M. Steup (Eds.), *A Companion to Epistemology* (pp. 93-105). Chicester: John Wiley & Sons, Ltd.

Hawthorne, J. (2004a). *Knowledge and Lotteries*. New York NY: Clarendon Press.

Hawthorne, J. (2004b). Three models of sequential belief updating on uncertain evidence. *Journal of Philosophical Logic, 33*, 89-123.

Hayne, H., & MacDonald, S. (2003). The socialization of autobiographical memory in children and adults: the roles of culture and gender. In R. Fivush & C. A. Haden (Eds.), *Autobiographical Memory and the Construction of a Narrative Self: Developmental and Cultural Perspectives*. Mahwah, N.J: L. Erlbaum.

Hempel, C. (1965). Aspects of scientific explanation *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (pp. 331-496). New York: Free Press.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*(1), 51-55. doi: 10.1111/j.1467-9280.1997.tb00543.x

Henrich, J., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., . . . Barr, A. (2005). "Economic man" in cross-cultural perspective: behavioral

experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*(6), 795-815. doi: 10.1017/S0140525X05000142

Hetherington, S. (1999). Knowing failably. *The Journal of Philosophy, 96*(11), 565-587.

Hoffman, P. (2002). Descartes's theory of distinction. *Philosophy and Phenomenological Research, 64*(1), 57-78. doi: 10.1111/j.1933-1592.2002.tb00142.x

Hollander, J. M., & Davies, T. F. (2009). Graves' Disease. In F. E. Wondisford & S. Radovick (Eds.), *Clinical Management of Thyroid Disease* (pp. 153-189). New York: Elsevier Health Sciences.

Huemer, M. (2011). Does probability theory refute coherentism? *The Journal of Philosophy, 108*(1), 35-54.

Hume, D. (1739a). Sect. vi. Of the inference from the impression to the Idea. *A Treatise of Human Nature, Book III* Retrieved 12 September, 2012, from http://ebooks.adelaide.edu.au/h/hume/david/h92t/B1.3.6.html

Hume, D. (1739b). Section iv Of the connexion or association of ideas. In eBooks@Adelaide (Ed.), A Treatise of Human Nature. Adelaide, S.A.: University of Adelaide. Retrieved from http://ebooks.adelaide.edu.au/h/hume/david/h92t/B1.1.4.html.

Hursthouse, R. (2013). Virtue ethics. *The Stanford Encyclopedia of Philosophy* Retrieved 6 September, 2013, from http://plato.stanford.edu/archives/fall2013/entries/ethics-virtue/

Jacoby, L. L. (1991). A process dissociation framework--separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541. doi: 10.1016/0749-596x(91)90025-f

James, W. (1890). The Principles of Psychology Retrieved 5 October, 2012, from http://ebooks.adelaide.edu.au/j/james/william/principles/index.html

Jeffrey, R. (1983). *The Logic of Decision* (2nd ed.). Chicago: University of Chicago Press.

Johnson, M. K. (2006). Memory and reality. *The American Psychologist, 61*(8), 760-771. doi: 10.1037/0003-066x.61.8.760

Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives, 24*(1), 281-323. doi: 10.1111/j.1520-8583.2010.00194.x

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological review, 103*(3), 582-591. doi: 10.1037/0033-295X.103.3.582

Kamil, A. C., & Gould, K. L. (2008). Memory in food caching animals. In J. H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (Vol. 1, pp. 419-439). Oxford, United Kingdom: Elsevier.

Kim, J. (1988). What is 'naturalized epistemology'? *Philosophical Perspectives, 2*, 281-405.

Klein, P. (1976). Knowledge, causality, and defeasibility. *The Journal of Philosophy, 68*(16), 471-482.

Klein, P., & Warfield, T. A. (1994). What price coherence? *Analysis, 54*(3), 129-132.

Knobe, J., & Nichols, S. (2008). *Experimental Philosophy*. New York: Oxford University Press.

Kornblith, H. (1985). Introduction: What is naturalistic epistemology? *Naturalizing Epistemology* (pp. 1-13). Cambridge, MA: The MIT Press.

Kornblith, H. (2002). *Knowledge and its place in nature*. Amazon Kindle Edition: Oxford Scholarship Online.

Kornblith, H. (2009). Sosa in perspective. *Philosophical Studies, 144*(1), 127-136.

Kornblith, H. (2012). *On reflection*. Oxford: Oxford University Press.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd enlarged with "Postscript-1969" ed.). Chicago: University of Chicago Press.

Laplace, P. (1820/1951). Essai Philosophique sur les Probabilités forming the introduction to his Théorie Analytique des Probabilités, Paris: V Courcier (F. W. Truscott & F. L. Emory, Trans.) *A Philosophical Essay on Probabilities*. New York: Dover.

Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science, 48*, 19-48.

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron, 73*(4), 653-676.

Leichtman, M. D., Wang, Q. W., & Pillemer, D. B. (2003). Cultural variations in interdependence and autobiographical memory: lessons from Korea, China, India and the United States. In R. Fivush & C. A. Haden (Eds.), *Autobiographical Memory and the Construction of a Narrative Self: Developmental and Cultural Perspectives*. Mahwah, N.J.: L. Erlbaum.

Lepock, C. (2010). A Virtue Epistemology: Apt Belief and Reflective Knowledge. [Book Review]. *Canadian Journal of Philosophy, 40*(1), 131-149.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological Review, 94*(4), 412.

Levine, D. N. (2007). Sherrington's "The Integrative Action of the Nervous System": a centennial appraisal. *Journal of the Neurological Sciences, 253*(1-2), 1-6. doi: 10.1016/j.jns.2006.12.002

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. LaSalle IL: Open Court.

Lewis, D. (1968). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Locke, J. (1689a). Chapter X. Of Retention. *An essay concerning human understanding* Retrieved 6 July, 2013, from http://ebooks.adelaide.edu.au/l/locke/john/l81u/B2.10.html

Locke, J. (1689b). Of the association of ideas *An essay concerning human understanding* Retrieved 6 July, 2013, from http://ebooks.adelaide.edu.au/l/locke/john/l81u/B2.33.html

Luminet, O., & Curci, A. (2008). *Flashbulb Memories: New Issues and New Perspectives*. Hoboken, NJ: Taylor & Francis.

Luria, A. R. (1987). *The Mind of a Mnemonist: A Little Book about a Vast Memory*. Cambridge MA: Harvard University Press.

Lycan, W. G. (1988). *Judgement and justification*. Cambridge, UK: Cambridge University Press.

Lyons, J. C. (2012). Should reliabilists be worried about demon worlds? *Philosophy and Phenomenological Research*, no-no. doi: 10.1111/j.1933-1592.2012.00614.x

Magnus, P. D., & Callender, C. (2004). Realist ennui and the base rate fallacy. *Philosophy of Science, 71*(3), 320-338.

Marcus, R. G. (1995). The anti-naturalism of some language centered accounts of belief. *Dialectica, 49*, 113-129.

Martin, C. B., & Deutscher, M. (1966). Remembering. *Philosophical Review, 75*, 161-196.

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (Vol. 4). Edinburgh, UK: Edinburgh University Press.

McEwen, B. S., & Wingfield, J. C. (2010). What's in a name? Integrating homeostasis, allostasis and stress. *Hormones and Behavior, 57*(2), 105.

Mill, J. (1878). Chapter III: The association of ideas *Analysis of the phenomena of the human mind* Retrieved 20 June, 2013, from http://archive.org/stream/analysisofthephe00milluoft - page/n151/mode/2up

Mill, J. S. (1882). Chapter IV. Of the laws of mind. In eBooks@Adelaide (Ed.), A System of Logic. Adelaide, S.A.: University of Adelaide. Retrieved from http://ebooks.adelaide.edu.au/m/mill/john_stuart/system_of_logic/chapter 59.html - chapter59.

Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge, MA: The MIT Press.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638-677.

Moore-Ede, M. C. (1986). Physiology of the circadian timing system: Predictive versus reactive homeostasis. *Am J. Physiol., 250* R737-R752.

Mouritsen, H., & Ritz, T. (2005). Magnetoreception and its use in bird navigation. *Current Opinion in Neurobiology, 15*(4), 406-414. doi: 10.1016/j.conb.2005.06.003

Mrosovsky, N. (1990). *Rheostasis: The Physiology of Change. .* Oxford, UK: Oxford University Press.

Musgrave, A. (1988). The ultimate argument for scientific realism. In R. Nola (Ed.), *Relativism and Realism in Science* (Vol. 6, pp. 229-252). Boston, MA: Kluwer Academic Publishers.

Nairne, J. S., & Pandeirada, J. N. S. (2008). 2.12 Forgetting. In H. B. John (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 179-194). Oxford, UK: Academic Press.

Nelson, R. J., & Drazen, D. L. (2000). Seasonal changes in stress responses. In G. Fink (Ed.), Encyclopedia of Stress (Vol. 3, pp. 402-408). San Diego, CA: Academic Press.

Neurath, O. (1983/1932). Protocol sentences. In R. S. Cohen & M. Neurath (Eds.), *Philosophical Papers 1913-1946*. Dordrecht: Reidel.

Nichols, S., & Stich, S. (2003). *Mindreading*. New York: Clarendon Press.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32*, 932-943.

Oaksford, M., & Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford, UK: Oxford University Press.

Olsson, E. J. (2010). Coherentism. In S. Bernecker & D. Pritchard (Eds.), *The Routledge Companion to Epistemology*. Oxford, UK: Routledge.

Olsson, E. J. (2012). Coherentist theories of epistemic justification. *The Stanford Encyclopedia of Philosophy* Winter, 2012. Retrieved 11 February, 2013, from http://plato.stanford.edu/archives/win2012/entries/justep-coherence/

Olsson, E. J., & Shogenji, T. (2004). Can we trust our memories? C.I. Lewis's Coherence Argument. *Synthese, 142*, 21-41.

Page, J. (2013, 9 July). British public wrong about nearly everything, survey shows, *The Independent*. Retrieved from http://www.independent.co.uk/news/uk/home-news/british-public-wrong-about-nearly-everything-survey-shows-8697821.html

Palmer, D. C. (2010). Operant conditioning. In I. B. Weiner & E. W. Craighead (Eds.), The Corsini Encyclopedia of Psychology and Behavioral Science. Hoboken, NJ: Wiley Online Library. Retrieved from

http://onlinelibrary.wiley.com/doi/10.1002/9780470479216.corpsy0623/abstract. doi: 10.1002/9780470479216.corpsy0623

Parmigiani, G., & Inoue, L. Y. T. (2009). Coherence *Decision Theory: Principles and Approaches* (pp. 13-31). Chichester, UK: John Wiley & Sons, Ltd.

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L.-W., Wamsley, E. J., Tucker, M. A., . . . Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory, 92*(3), 327-334. doi: 10.1016/j.nlm.2009.03.007

Pearl, J. (1985, 15-17 August). *Bayesian networks: a model of self-activated memory for evidential reasoning.* Paper presented at the Cognitive Science Society, UC Irvine.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Petroski, H. (1985). *To Engineer is Human: The Role of Failure in Successful Design*. New York, NY: St. Martin's Press.

Pfeffer, A., & Tai, T. (2012). Asynchronous dynamic Bayesian networks Retrieved 11 April, 2013, from http://arxiv.org/abs/1207.1398

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Plantinga, A. (1993). Bonjourian coherentism *Warrant: The Current Debate* (pp. 87-113). New York: Oxford University Press.

Plato. (2012). Theaetetus. *eBooks@Adelaide* Retrieved 1 August, 2013, from http://ebooks.adelaide.edu.au/p/plato/p71th/theaetetus.html - theaetetus

Polyn, S. M., Kahana, M. J., & Norman, K. A. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129-156. doi: 10.1037/a0014420

Price, H. H., Laird, J., & Wright, J. N. (1936). Symposium: Memory-knowledge. *Proceedings of the Aristotelian Society, Supplementary, 15*, 16-60.

Prince, S. E., Tsukiura, T., & Cabeza, R. (2007). Distinguishing the neural correlates of episodic memory encoding and semantic memory retrieval. *Psychological Science, 18*(2), 144-151. doi: 10.1111/j.1467-9280.2007.01864.x

Pritchard, D. (2009). Apt performance and epistemic value. *Philosophical Studies, 143*(3), 407-416. doi: 10.1007/s11098-009-9340-7

Pritchard, D., & Turri, J. (2012). The value of knowledge. *The Stanford Encyclopedia of Philosophy* Winter 2012. Retrieved 7 May, 2013, from http://plato.stanford.edu/archives/win2012/entries/knowledge-value/

Putnam, H. (1978). *Meaning and the Moral Sciences*. London, UK: Routledge & Kegan Paul.

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

Quine, W. V., & Ullian, J. S. (1970). *The Web of Belief*. New York: Random House.

Raibert, M., Blankespoor, K., Nelson, G., Playter, R., & Boston Dynamics. (2013). BigDog, the rough-terrain quaduped robot Retrieved 14 April, 2013, from http://www.bostondynamics.com/img/BigDog_IFAC_Apr-8-2008.pdf

Ramsey, P. F. (1964). Truth and probability. In H. E. Kyburg & H. E. Smokler (Eds.), *Studies in Subjective Probability*. New York: Wiley.

Reid, A., & Gaskell, G. (2012). The effects of sleep and time on semantic memory stabilisation and integration. *Journal of Sleep Research 21*, 248-248.

Reid, T. (1785). Memory. *Essays on the intellectual powers of man* Retrieved 14 November, 2012, from http://www.earlymoderntexts.com/pdf/reidint3.pdf

Rescorla, M. (2011). Convention. *The Stanford Encyclopedia of Philosophy* Spring 2011. Retrieved 27 March, 2013, from http://plato.stanford.edu/archives/spr2011/entries/convention/

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: Some foundational issues. *Learning and Individual Differences, 7*, 145-162. doi: 10.1026/1041-6080(95)90028-4

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron, 61*(2), 168-185. doi: http://dx.doi.org/10.1016/j.neuron.2009.01.002

Richter, C. P. (1943). Total self-regulatory functions in animals and human beings. *Harvey Lectures, 38*, 63-103.

Roche, W. (2010). Coherentism, truth, and witness agreement. *Acta Analytica, 25*(2), 243-257. doi: 10.1007/s12136-009-0065-1

Rogers, K. (Ed.). (2012). *The Endocrine System*. New York: Britannica Educational Publications.

Romero, L. M., Dickens, M. J., & Cyr, N. E. (2009). The reactive scope model—a new model integrating homeostasis, allostasis, and stress. *Hormones and Behavior, 55*(3), 375-389.

Rosenbaum, R. S., Murphy, K. J., & Rich, J. B. (2012). The amnesias. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(1), 47-63. doi: 10.1002/wcs.155

Rubin, D. C., & Greenberg, D. L. (1998). Visual memory-deficit amnesia: A distinct amnesic presentation and etiology. *Proceedings of the National Academy of Sciences of the United States of America, 95*(9), 5413-5416. doi: 10.1073/pnas.95.9.5413

Russell, B. (1921). *The Analysis of Mind*. London, UK: General Books LLC.

Russell, B. (2013). A priori justification and knowledge. *The Stanford Encyclopedia of Philosophy* Summer 2013. Retrieved 2 September, 2013, from http://plato.stanford.edu/archives/sum2013/entries/apriori/

Ryle, G. (1946/1971). Knowing how and knowing that *Gilbert Ryle: Collected Papers* (Vol. 2, pp. 212-225). New York, NY: Barnes and Noble.

Ryle, G. (1949). *The Concept of Mind*. Chicago: Chicago University Press.

Salmon, W. (1971). Statistical explanation *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology, 13*, 501-518.

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2012). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*. doi: 10.1037/a0030777

Schulkin, J. (2003). *Rethinking homeostasis: Allostatic regulation in physiology and pathophysiology*. Cambridge, MA: The MIT Press.

Schwitzgebel, E. (2011). Belief. *The Stanford Encyclopedia of Philosophy* Winter 2011. Retrieved 5 July, 2013, from http://plato.stanford.edu/archives/win2011/entries/belief/

Scoville, W. B. (1968). Amnesia after bilateral mesial temporal-lobe excision: Introduction to case H.M. *Neuropsychologia, 6*(3), 211-213. doi: 10.1016/0028-3932(68)90020-1

Shah, H. (2013). The perils of perception: Tackling the divide between public views and the evidence Retrieved 18 July, 2013, from http://www.rssenews.org.uk/2013/07/rss-commission-new-research-into-public-perceptions-of-statistics/

Shanahan, M. (2009). The Frame Problem. *The Stanford Encyclopedia of Philosophy* Winter 2009. Retrieved 27 June, 2013, from http://plato.stanford.edu/archives/win2009/entries/frame-problem/

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379-423.

Sherrington, C. (1941). *Man on His Nature*. New York: the Macmillan Company.

Sherrington, C. S. (1906). The Integrative Action of the Nervous System Retrieved from http://ia600303.us.archive.org/24/items/integrativeactio00sheruoft/integrativeactio00sheruoft.pdf

Siebel, M., & Wolff, W. (2008). Equivalent testimonies as a touchstone of coherence measures. *Synthese, 161*(2), 167-182. doi: 10.1007/s11229-006-9155-4

Slack, K. E., Jones, C. M., Ando, T., Harrison, G. L., Fordyce, R. E., Arnason, U., & Penny, D. (2006). Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Molecular Biology and Evolution, 23*(6), 1144-1155. doi: 10.1093/molbev/msj124

Slobodchikoff, C. N., Perla, B. S., & Verdolin, J. L. (2009). *Prairie Dogs: Communication and Community in an animal society*. Cambridge, MA: Harvard University Press.

Sorabji, R. (2004). *Aristotle on Memory* (2nd ed.). London: Gerald Duckworth & Co.

Sosa, E. (1980). The raft and the pyramid: Coherence versus foundations in the theory of knowledge. *Midwest studies in philosophy, 5*(1), 3-26.

Sosa, E. (1993). Proper functionalism and virtue epistemology. *Noûs, 27*(1), 51-65.

Sosa, E. (1997). Reflective knowledge in the best circles. *Journal of Philosophy, 94*(8), 410-430.

Sosa, E. (1999). How to defeat opposition to Moore. *Noûs, 33*(s13), 141-153. doi: 10.1111/0029-4624.33.s13.7

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Oxford, UK: Oxford University Press.

Sosa, E. (2009a). *Reflective Knowledge: Apt Belief and Reflective Knowledge* (Vol. 2). Oxford, UK: Oxford University Press.

Sosa, E. (2009b). Replies to commentators on A Virtue Epistemology (Oxford University Press, 2007). *Philosophical Studies, 144*(1), 137-147. doi: 10.1007/s11098-009-9374-x

Sosa, E. (2011). *Knowing Full Well*. Princeton, NJ: Princeton University Press.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89-96. doi: 10.1111/j.1467-7687.2007.00569.x

Spiegel, D. (2012). Divided consciousness: dissociation in DSM-5. *Depression and Anxiety, 29*(8), 667-670. doi: 10.1002/da.21984

Stanford, P. K. (2010). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.

Stanley, J. (2011). Knowing (How). *Noûs, 45*(2), 207-238. doi: 10.1111/j.1468-0068.2010.00758.x

Stanovich, K. E. (1999). *Who is Rational?: Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E., & West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral & Brain Sciences, 23*, 645-726.

Starling, E. H. (1923). *The Wisdom of the Body*. London: H. K. Lewis.

Stein, E. (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. Oxford, UK: Clarendon.

Suddendorf, T., & Moore, C. (2011). Introduction to the special issue: The development of episodic foresight. *Cognitive Development, 26*(4), 295-298. doi: 10.1016/j.cogdev.2011.09.001

Suppes, P. (1970). A Probabilistic Theory of Causality Retrieved 25 June, 2013, from http://suppes-corpus.stanford.edu/article.html?id=106-1

Sutton, J. (2006). Introduction: memory, embodied cognition, and the extended mind. *Philosophical Psychology, 19*(3), 281-289. doi: 10.1080/09515080600702550

Talbott, W. (2011). Bayesian epistemology. *The Stanford Encyclopedia of Philosophy* Retrieved 2 January, 2013, from http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/

Tarricone, P. (2011). *The Taxonomy of Metacognition*. New York: Psychology Press.

Thelen, E., & Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: The MIT press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189-208. doi: 10.1037/h0061626

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Oranization of memory* (pp. 381-403). New York: Academic Press.

Turri, J. (2012). Is knowledge justified true belief? *Synthese, 184*(3), 247-259. doi: 10.1007/s11229-010-9773-8

Van Cleve, J. (2011). Can coherence generate warrant ex nihilo? Probability and the logic of concurring witnesses. *Philosophy and Phenomenological Research, 82*(2), 337-380. doi: 10.1111/j.1933-1592.2010.00487.x

Van Valkenburgh, B. (2007). Déjà vu: The evolution of feeding morphologies in the Carnivora. *Integrative and Comparative Biology, 47*(1), 147-163. doi: 10.1093/icb/icm016

von der Malsburg, C., Phillips, W. A., & Singer, W. (Eds.). (2010). *Dynamic Coordination in the Brain: From Neurons to Mind (Strüngmann Forum Reports)*. Cambridge, MA: The MIT Press.

Wagner, C. (2013). Is conditioning really incompatible with holism? *Journal of Philosophical Logic, 42*(2), 409-414. doi: 10.1007/s10992-012-9230-x

Weinberg, J. M., Nichols, S., & Stich, S. (2008). Normativity and epistemic intuitions. *Experimental Philosophy, 29*, 17-45.

Weiner, I. (2012). *Handbook of Psychology: History of Psychology*. Chichester, NY: Wiley.

Weisberg, J. (2009). Commutativity or holism? A dilemma for conditionalizers. *The British Journal for the Philosophy of Science, 60*(4), 793-812. doi: 10.1093/bjps/axp007

Weiskopf, D. A. (2008). Patrolling the mind's boundaries. *Erkenntnis, 68*(2), 265-276. doi: 10.1007/s10670-007-9095-5

Wheeler, G., & Scheines, R. (2013). Coherence and confirmation through causation. *Mind, 122*, 135-170.

Williamson, J. (2013). Why frequentists and Bayesians need each other. *Erkenntnis, 78*(2), 293-318. doi: 10.1007/s10670-011-9317-8

Wingfield, J. C., & Ramenofsky, M. (1999). Hormones and the behavioural ecology of stress In P. H. M. Balm (Ed.), *Stress Physiology in Animals*. Sheffield, UK: Sheffield Academic Press.

Worrall, J. (2007). Miracles, pessimism and scientific realism. *British Journal for the Philosophy of Science*, 151-152.

Xu, F., Dewar, K., & Perfors, A. (2009). Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism. *The Origins of Object Knowledge*, 263-284.

Xu, Y., & Wang, P. (2012). The frame problem, the relevance problem, and a package solution to both. *Synthese, 187*(1), 43-72. doi: 10.1007/s11229-012-0117-8

Yataco Md, A. R., Fleisher Md, L. A., & Katzel Md, P. L. I. (1997). Heart rate variability and cardiovascular fitness in senior athletes. *The American Journal of Cardiology, 80*(10), 1389-1391. doi: http://dx.doi.org/10.1016/S0002-9149(97)00697-8

Young, G. (2008). Clarifying "familiarity": Examining differences in the phenomenal experiences of patients suffering from prosopagnosia and capgras delusion. *Philosophy, Psychiatry, & Psychology, 14*(1), 29-37.

Zagzebski, L. (1996). *Virtues of the Mind*. Cambridge, UK: Cambridge University Press.

Zhao, J., & Osherson, D. (2010). Updating beliefs in light of uncertain evidence: Descriptive assessment of Jeffrey's rule. *Thinking & Reasoning, 16*(4), 288-307. doi: 10.1080/13546783.2010.521695