Four Paradigms in Comparative Psychology

by

David Evan Pence

B.A. in Philosophy, Ohio University, 2013

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

David Evan Pence

It was defended on

June 2, 2020

and approved by

Colin Allen, Distinguished Professor, History and Philosophy of Science

Mazviita Chirimuuta, Associate Professor, History and Philosophy of Science

James Lennox, Emeritus Professor, History and Philosophy of Science

Wayne Wu, Associate Professor, The Center for the Neural Basis of Cognition

Dissertation Director: Edouard Machery, Distinguished Professor, History and Philosophy of Science

Four Paradigms in Comparative Psychology

David Evan Pence, PhD

University of Pittsburgh, 2020

The following examines the development of comparative psychology and the evidence, arguments, and epistemological challenges that have characterized its approach to the question of animal rationality. I distinguish between four modes of research that come to prominence at different points in its history, the natural historical, strict behavioral, cognitive, and neurophysiological, analyzing each through a critical episode in its development and the set of claims associated with the approach. The first study concerns the field's Darwinian origins and its early commitment to the fundamental similarity of human and animal minds. I argue from a close reading of Darwin's notebooks that the critical break for the nascent field came not from an antecedent endorsement to evolutionary theory, as commonly supposed, but a set of political and philosophical commitments inherited from the Enlightenment. Next, I show how this approach proved vulnerable to attack from younger and more positivistic psychologists in the twentieth century. I analyze why the Darwinians were accused of employing less than scientific methods, explaining how this fact helped precipitate a shift toward more conservative standards of evidence and strictly lab-based research. From there, I consider how the behavioral tools of this era have left modern 'cognitive' research with nagging underdetermination issues. I argue that strictly behavioral methods cannot tell us what the nature of animal thought is but that other methods may. Finally, I consider the state of the rationality debate at present. Drawing on the most recent evidence from systems neuroscience, I argue that animals as distant as rats have the capacity to engage in basic forms of reasoning ventured by Darwin and suspected but never quite shown in the cognitive era.

Table of Contents

Preface ix

1.0 First Paradigm: Darwin and the Origins of the Mental Continuity Thesis 1

  1.1 Introduction 1

  1.2 The Puzzle 3

  1.3 “Thy brother Emmets, and thy sister Worms!” 12

  1.4 Enter Darwin 21

    1.4.1 Early Life 22

    1.4.2 “One thinking sensible principle” 26

  1.5 Making Sense of Darwin’s Psychology 33

2.0 Second Paradigm: How Comparative Psychology Lost its Soul 37

  2.1 Introduction: A Tone Shift 37

  2.2 Psychology Old and New 40

  2.3 Historical Context: Institutional and Cultural Shifts 46

  2.4 Animals and Psychical Research 51

  2.5 Two Lines of Attack 56

    2.5.1 The Methodological Critique 57

    2.5.2 Cementing the Approach 64

  2.6 Conclusion: Exorcizing Ghosts 69

3.0 Third Paradigm: Cognitivism and the Case of the Cognitive Map 71

  3.1 Introduction 71

  3.2 The Cognitive Challenge 73

3.3 A Missing Factor 78
    3.3.1 Dimensions of Measurement 79
    3.3.2 Dimensions Applied to Animal Cognition 83
3.4 Case Study: The Cognitive Map 88
    3.4.1 Outlines of the Controversy 89
    3.4.2 Explaining the Controversy 92
3.5 The Cognitive Dilemma 97
4.0 Fourth Paradigm: Building a Neurobiological Case for Animal Rationality 100
4.1 Introduction: Dog Logicke 100
4.2 Reason and its Evidence 102
    4.2.1 Isolating the Phenomenon 102
    4.2.2 Making a Neurobiological Case 108
4.3 Executive Capacities and the Prefrontal Cortex 110
    4.3.1 Computational Underpinnings 111
    4.3.2 Evidence 114
    4.3.3 Objections 120
4.4 Imagination and the Hippocampus 124
    4.4.1 Principles of Prospective Imagination 126
    4.4.2 Evidence 128
    4.4.3 Objections 134
4.5 Conclusion: Rational Animals 139
5.0 Postscript: Open Questions 141
5.1 Moral Dimensions 141

5.2 Evolution and Function 145

5.3 Human Uniqueness 147

Bibliography 150

List of Figures

Figure 1 The New Morality **20**

Figure 2 Bob the Telepathic Dog **54**

Figure 3 Thorndike Puzzle **59**

Preface

When I first started this project, I was struck by just how much I came across animal cognition in everyday life. Having worked primarily in the philosophy of perception, I'd seen the occasional news story or social media post on my areas of interest. A significant study or ongoing debate might be picked up by Reuters or covered in popular science outlets like *Nautilus*. In general, though, the discussion took place largely between devotees. Explaining the workings of perceptual inversion experiments or sensory substitution devices to family and friends required a fair amount of unpacking and rarely elicited strong opinions. There was certainly interest, but engagement proved less common. Working on animal minds is different. Since starting the work, hardly a week has gone by without my seeing something in the media—lab rats learning to drive tiny cars, say, or Youtube videos of cats bewildered at seeing caretakers' faces change under a snapchat filter. In explaining what I do to non-specialists, I've found my conversation partners far more forthcoming with views and anecdotes. People are interested in other animals,[1] in what they're thinking, and whether they are as "like us" as they frequently seem.

That we are so interested isn't a feature limited to our present age. References to the cognitive capacities of our fellow creatures can be traced at least as far back as the sixth century BCE, when the Aitaraya Upanishad ventured that animals neither "say what they have known" nor "know what is to happen to-morrow" (1879, p. 5). Moving forward in time, moreover, one finds discussion in nearly every century since. The animal mind and its relation to the human appears in

---

[1] "Animals," "other animals," and "non-human animals" shall be used interchangeably.

poetry, legal codes, religious texts, and philosophical discourses. It can be found in the *Edicts* of Ashoka and the *Metamorphoses* of Ovid, the *Psychology* of Avicenna and Aristotle's *History of Animals*. In the couple of centuries on which I focus (roughly the early 19th century onward), it has been a central preoccupation of the professional sciences as well. In each of these domains and for most of this time, it has also proven a heated topic. Both defenders and opponents of animal rationality and the thesis of human-animal mental continuity have been labelled fools, sophists, and brutes by their interlocutors. (Of all the conferences I've attended, it has been those on animals that have initiated the most passioned debate). There are, I suspect, at least three reasons for this sustained interest.

The first comes down to metaphysical stakes. For better or worse, there is a lot tied up in the question of animal rationality. In many intellectual and cultural traditions, including those most influential in the shaping of the present-day science of animal minds, the ability to foresee and deliberate has been used to set humanity apart. Humans, as figures from Avicenna to Davidson have told us, are unique in our capacity to reason and comprehend; we alone are able to see the future and guide our actions in accordance with our interests and ends. It's a claim so obvious to some as made without evidence (even among defenders of animal interests; e.g., Korsgaard, 2004). Yet, it is also a claim that many deem essential. Reason is what gives us our collective aura of dignity. Far more than being *the* animal that wears clothes or *the* flat-nailed featherless biped, as Plato supposedly defined us, our being *the* rational animal is what makes us special. In some traditions, it even imparts to us a certain semblance to God, characterized as a supreme intelligence (and depicted only conventionally as bipedal or clothes-wearing). Were this status lost, some fear, it would threaten our sense of place. Allowing that our most prized possession had analogues in apes, dogs, or rats could mean the "brutalization and degradation" the species (to use T.H.

Huxley's phrase). Historically, defending human uniqueness has been seen as a way of defending humanity itself.

Second and relatedly, one's answer to the question has a certain ethical significance. Rationality is widely thought to bring with it an intrinsic moral worth, and a major shift in its recognized distribution would matter to how we conduct ourselves (Carruthers, 1992; Korsgaard, 2004). The notion is frequently tied to ideas of moral personhood, for instance. An organism lacking the capacity to recognize its own interests or value its continued existence, it's argued, has little demand on our doing so (Tooley, 1972; Singer, 1979). One that is a person, by contrast, may be deserving of certain legal protections and privileges. Such a case has recently been made by the Nonhuman Rights Project in pursuing legal personhood for Chimpanzees (Andrews et al., 2018). What's more, it would seem to increase its holder's capacity for suffering and flourishing both, bringing with it a new set of moral considerations about their treatment and housing (DeGrazia, 2002). When ethical guidelines for animal research have been revised, as in the case of primates (e.g., APC, 2002), the subjects' cognitive faculties and their similarity to humans have consistently numbered among the reasons cited. Were we to learn that animals like pigs, whom we eat, or rats, whom we trap and perform our most invasive experimental procedures upon, were endowed with the same basic capacity to judge or reflect upon their future, past, and present situation, it would surely give us pause.

Finally, there's the intrinsic interest and challenge of the question. Living with animals, one is constantly faced with our shared communicative limitations. As Dennett (1996, ch. 1) observes, our principal means of finding out about the thought processes of other humans is linguistic. In day to day life, if we have any doubts about what someone is thinking or how they have arrived at some behavior, we can simply ask them; even the most positivistic versions of human psychology build on the ability to explain task demands and elicit feedback from subjects. When it comes to other animals, however, these tools are of limited value. The *Upanishads* were right to say that animals cannot say what they have known, at least not in much detail or in a manner comprehensible to us. Without the guidance of a shared language, the mind naturally wanders. Seeing the actions of a household cat or dog, one fills in what is left unsaid by appeal to thoughts and interests. Evidentially, the situation is rather like reading a novel in fragments, and just as in the case of textual interpretation, the ambiguities soon give rise to competing schools. What is said by one to reflect a process of reasoning can, as centuries of disagreement show, often be dismissed by another as the product of habit or instinct. Questions of interpretation thus lead to questions of method—of how heavily to weight simplicity in one's explanations, whether to consider moral implications in assessing claims, and so forth. The desire to know comes always to the need to justify and convince.

The era on which I focus is not so broad as the foregoing discussion might suggest. All told, the dissertation covers only that fragment of the animal minds debate found in modern comparative psychology (a science whose origins lie, as I noted earlier, somewhere in the 19th century). As I see it, however, the work is oriented toward the same fundamental issues and concerns. In researching and writing it, I hoped to understand how the religious and moral stakes of the topic weighed on the discussion, to uncover when and why certain standards and forms of evidence came to dominate its discussion, and to determine what evidence, if any, could settle the matter. Beyond the question of whether animals reason—something addressed only in the final chapter—I wanted to see how, in the era of "scientific" inquiry, we've come to believe or doubt that they do. The result is an episodic narrative arranged, as the title suggests, around four research paradigms: the natural historical, the behavioral, the cognitive, and the nascent neurophysiological.

The first two approaches are largely historical, finding few adherents in the present day; the third represents the most common contemporary approach, as judged by standard textbooks and articles; and the last reflects my own sense of where the field is going. The way each is discussed and how these discussions are distributed across chapters reflects these differences. The first two chapters are focused on the field's development: why it adopted some standards rather than others, how surrounding context impacted its conceptualization of other animals, and so forth. The chapters concerning present and future approaches, by contrast, are examined with more forward-looking and normatively loaded concerns in mind: what kind of evidence is needed to establish claims of reasoning in animals, does this evidence vindicate or undermine the old stories of animals pausing and deliberating over their choices, and the like. Taken as a whole, they make visible a familiar narrative of birth, exile, and return. Comparative psychology begins with a robust sense of reasoning in animals, loses it, and with much difficulty finds its way back, though much changed by the journey.

The work begins with Darwin, whose work comparing human and animal reasoning, emotion, and instinct in the 1870s is widely understood as the beginning of the field. Following the wild notoriety of his *Origin of Species*, Darwin turned—as the close of that text hinted—to the task of putting psychology on a new foundation. Bringing a diverse range of observations, anatomical reports, and field experiments to bear on the question, the naturalist made the case that human and non-human minds were fundamentally similar and demonstrated for those who would follow how these commensurable qualities could be put one against the another. The work, perhaps the most controversial of his career, would serve as an inspiration for researchers like George Romanes, Lauder Lindsay, and Wesley Mills in the coming decades. Yet the reasons and motivations behind Darwin's arguments—and the birth of the discipline—remain somewhat unclear. Some trace the work's beginnings to his thoughts on natural selection, others cite his readings in the sensationalist philosophy of Hume and Locke as key. My aim in chapter one is to present an alternative, if overlapping origin story. The driving force behind Darwin (and many of his followers), I argue, wasn't a theory of speciation or epistemological stance so much as a moral vision. Examining his upbringing and notebooks, I show how Darwin drew ideas and commitments from a range of animal welfare advocates, including poets like Shelley, philosophers like Voltaire and Priestley, and his own grandfather Erasmus. I review how these influences manifest in the formation of his views in the 1830s and why they stood out from Victorian society as a whole. In doing so, I argue for a rethinking of comparative psychology's start and its relation to questions of value.

From this starting point of sympathy and kinship, I then consider how the field moved away from its Darwinian tendencies and toward a far more conservative, lab-based, and human-exceptionalist approach. Though commonly presented as a matter of straightforward scientific progress—or even as the birth of the field qua science—I argue that the change had less to do with the growth of experiment or the introduction of reasonable standards of evidence than with the underlying ideology and moral vision of the field. In particular, I claim that the sympathetic vision of human-animal relations that underwrote the discipline's early decades rendered it vulnerable to attack with the rise of occult or 'psychical' research, a rapidly growing field championed by psychologists like William James and eminent animal researchers like George Romanes. Examining the writings of contemporary psychologists and naturalists, I show how the personal and methodological affinities between animal and psychical researchers—their shared emphasis on sympathy for one's subjects, use of field observation and anecdote, association with political reform movements, etc.—put the former at a rhetorical disadvantage and how this was used to great effect by early behaviorally minded researchers. If correct, the strict standards of evidence, strong preference for lab work over naturalistic observation, and general hostility to the idea that animals possess human-like mental faculties that have long characterized the field may be little more than accidents of history.

Chapter three picks up a few decades later, when the storied "cognitive revolution" began to introduce—or re-introduce—many of the traditional, mentalistic questions that had been rejected as anthropomorphic by their strict behavioral predecessors. The animal psychology that emerged in the middle of the twentieth century and continues to dominate the field allowed researchers to work on empathy, numerical reasoning, and many other areas that would have raised the specter of anthropomorphism in previous decades. In approaching them, however, the field nevertheless kept hold of many of the methodological precepts of the earlier paradigm. This combination of behavioral tools and mental questions has, in turn, been associated with a set of well-known underdetermination problems. The field has been at pains to resolve these issues but, I argue, is unlikely to do so as long as it continues to investigate cognitive questions with behavioral tools. Using concepts from the epistemology of measurement and the neurobiology of action control, I explain how the mismatch emerges and why alternative solutions to the problem are unlikely to help. I then illustrate perhaps the best-known controversy in the cognitive paradigm's history: namely, whether rats make use of 'cognitive maps' in memory and navigation. Should the field wish to avoid these problems in the future, it will need to relax its evidential standards, move away from purely behavioral research, or both.

Finally, as the work comes to a close, I return to the question with which we started, viz. whether other animals really do reason. Reviewing the evidence and maintaining to the best of my ability those scruples entrenched in the discipline itself, I find that we must ultimately side with

Darwin. Animals are indeed rational, insofar as they are capable of systematically organizing their thoughts, simulating potential actions, and evaluating their outcomes. Each, I argue, can be shown by scrutinizing the patterns of neural activity observed in regions homologous to those serving these functions in humans. Strong evidence for a basic grasp of ordering and systematic cognitive control, for instance, may be found in the dorsolateral prefrontal cortex of macaques, a region that, when damaged in humans leads to profound deficits in rule-following and task organization. Integrating this evidence, I argue that Darwin was ultimately correct in believing that non-human animals engage in deliberation. In closing, I consider what the results of this inquiry mean for our conception of rationality, our relation to other members of the animal kingdom, and the responsibilities that we have to them, suggesting a return to some of the older sentiments as we return to the older position on reason.

1.0 First Paradigm: Darwin, the Enlightenment, and the Origins of Human-Animal Mental Continuity

1.1 Introduction

Like most sciences, it is difficult to give a precise starting point for comparative psychology. In theory, one could extend the practice back to medieval bestiaries or ancient natural histories. Within the field, however, credit is usually awarded to Charles Darwin, whose *Descent of Man* and *Expression of Emotions in Man and Animals* are thought to mark the first major works in the science of animal minds, and it seems fitting that we should begin our study with him. Standard histories bear names like *From Darwin to Behaviorism* (Boakes, 1984) and laud the naturalist as “the earliest, and still most significant, figure” in its century and a half of existence (Innis, 1998, p. 3). In his willingness to draw parallels between human and animal minds, it’s said, Darwin “launched modern comparative and animal psychology” (Henley, 2017, p. 281). For all the praise heaped upon the naturalist, however, the reasons for Darwin’s commitment to mental continuity have remained somewhat mysterious. Disciplinary histories, though quick to cite Darwin’s psychological works, rarely dwell on their development. His central contention that there exists “no fundamental difference between man and the higher mammals in their mental faculties” (1871, p. 35) has long sat poorly with the field, leaving many only too eager to bypass what’s been labeled “Darwin’s Mistake” (Penn et al., 2008). Meanwhile, the majority of Darwin scholarship has—understandably—focused on speciation, inheritance, and other topics that are today grouped under the heading of biology. Works dedicated to his psychology remain comparatively rare, leaving us with few accounts of how his distinctive and critical commitment to mental continuity might have developed.[2] The field's origins are, in this sense, a bit of a mystery.

At present, there are two main hypotheses regarding the origin of Darwin's psychological views. Neither, I argue, is ultimately sufficient. The first, suggested by the presentation of the material in the *Descent*, sees his strident defense of mental continuity as an application of the ideas he'd developed in his evolutionary thinking (e.g., Swisher, 1967). Having arrived at a gradualist theory of evolution for features like beaks, the thinking goes, it was only a matter of time before Darwin or someone else extended the argument to thought processes. The second, developed at length in now-classic work by Richards (1987), argues that these conclusions were reached instead by Darwin's readings in the sensationalist tradition of British and French psychology. On this view, the naturalist's belief in the commensurability of human and animal minds had key precedents in figures like Locke, Hume, and Darwin's own grandfather, Erasmus. It's fair to say that these hypotheses contribute important aspects to our understanding of Darwin's psychology, yet they ultimately leave too many questions unanswered. Darwin's psychological views are too idiosyncratic and appear too early in his development for them to have arisen from either.

In the following, I argue that the key to understanding the development of Darwin's psychological thought and hence the origins of the comparative project lies in the moral culture from which it emerged. Darwin's defense of mental continuity has its roots less in the abstract philosophical grounds of sensationalism or transmutationism than the morality of his Whig Dissenter surroundings. More specifically, I claim, it can be traced to his upbringing and continued participation in a community with an overriding interest in animal welfare, a pronounced

---

[2] Richards (1987) and Gruber's (1974) monographs represent outstanding exceptions to this trend.

skepticism toward theological exceptionalism, and a relative hostility toward several the era's received hierarchies (ecclesiastical, aristocratic, etc.). Darwin's family and upbringing were as radical as any naturalist of the time, while the notes he kept while forming his core ideas are interlaced with disbelief and exasperation at the status of animals in the works of his fellow naturalists and Victorian society generally. The mental kinship of humans and animals, I claim, wasn't something Darwin had to be convinced of but something he felt from the beginning.

1.2 The Puzzle

The first thing one needs to understand about Darwin's views on human and animal minds is its reach. The human and animal mind were not only of the same origin, the naturalist contended, but of the same fundamental kind. Although great differences of degree obtained, and while humans had quite a few "peculiar" capacities, there existed no dimension of human mentality for which a counterpart in creatures like apes and dogs couldn't be found. Curiosity, imagination, self-awareness, reason, and even morality existed among the higher mammals. On occasion, they could be observed in birds and insects, as well.[3] His expressed reasons for the belief were both diverse and expansive. Darwin called upon traveler's reports, personal observations, comparative anatomy, and common knowledge among the reading public. Humans and animals, he noted, bore striking resemblance in organs of sense and of mind, suffered the same pathologies and affections, and displayed manifold behaviors indicative of complex inner lives. The reader is treated to vivid

---

[3] Darwin's approaches certain plants in a similar manner (see below).

descriptions of the human in our animal relatives, ranging from the exalted love of a dog for its pups to the laughable low of baboons recoiling after a night of heavy drinking (p. 12).

When arguing for many of these capacities, it seems, Darwin met with comparatively little resistance. As Richards (1987, p. 197) notes, much of the *Descent*'s force came from its author's use of premises readers were perfectly willing to grant. Few Englishmen would besmirch the loyalty of a dog or deny that nature had furnished the animal kingdom with maternal instincts and affections. Amid the panoply of mental characteristics discussed in the book's early chapters, however, a certain class stood out. Beyond the light-hearted discussions of play and affection, Darwin believed that animals possessed reason, with all its philosophical and theological implications. Even those "extremely low in the scale," he claimed, attained some measure of the faculty both he and his audience believed to be "the summit" of the human mind (1871, p. 46). When combined with instinct, memory, and imagination, he added, still more striking forms emerged. Abstraction, self-awareness, language, and tool use could all be observed by a sufficiently attentive naturalist. Perhaps most disconcerting for his Victorian audience, something like moral agency could too. A dog of good memory might well reminisce in a simple form of self-consciousness (p. 58; p. 62), and wedding such capacities to certain social instincts, might experience pangs of conscience for past wrongs (p. 78). At one point, Darwin even suggests that a simple spiritual life might emerge. A monkey or hound accustomed to a human caretaker, for instance, might develop some religious devotion or infer from the depth of the human's powers that they represent a higher form of being (p. 68).

This was a bridge too far. While the idea that other animals had human-like passions or thoughts grounded in immediate sensation wasn't terribly troubling, the thought that they may be capable of abstraction, self-awareness, or conscience proved unacceptable. Looking to the press coverage of the time and the copious letters received after his publication of the *Descent*—and even before, when the parallels had been only hinted at—one finds alternatingly cool or harsh reception. The topic of human exceptionalism had been the subject of heated exchange with friends like Lyell and Wallace in the years prior to the work's publication,[4][5][6] while its release prompted a wave of skeptical comment (Browne, 2002, pp. 350-352). In an anonymously published review in the *Quarterly Review,* for instance, Mivart accused Darwin of more than one "groundless inference" regarding reason (1871, p. 74), describing his interpretations as "thoroughly gratuitous" (p. 69) and unscientific. As a rational, free, and moral agent, he claimed "man…differs far more from an elephant or a gorilla than do these from the dust of the earth on which they tread," and in denying this, "the author of the 'Descent of Man' has utterly failed in the only part of his work which is really important" (p. 89). A review in the *Times* went so far as to claim that Darwin's ideas might unleash "murderous revolutions" on grounds of its eliminating the essential difference on which society depended: "At a moment when every artificial principle of authority seems undermined," the reviewer intoned, "we have no other guarantee for the order and peace of life except in the eternal authority of those elementary principles of duty which are independent of all times and of all circumstances." The divide between Darwin and the received wisdom of most of his countrymen prompts a question: Why? What is it about Darwin's background that inclined him toward such a controversial view?

In the history of thinking about Darwin's views on Man, there have been a few hypotheses offered in response. One natural thought is that they followed from his groundbreaking work on

---

[4] Darwin to Lyell, 9 October, 1866, letter 5236

[5] Darwin to Wallace, 27 March, [1869], letter 6684

[6] Darwin to Wallace, 14 April, 1869, letter 6706

evolution by natural selection (see, e.g., Swisher, 1967). The unification of human and animal origins contained in Darwin's writings on transmutation, the thinking goes, found a second home in his unification of human and animal psychology. It's an idea suggested by Darwin as much as anyone. The *Origin*'s penultimate sentence indicates that that "psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation" (1859), while the opening of his *Descent*, published more than a decade later, presented the work as an effort "to see how far the general conclusions arrived at in my former works were applicable to man" (1871, p. 2). In the latter text, it may be added, Darwin often addresses his remarks directly to critics of the *Origin* who cited humans' psychological capacities as evidence against his evolutionary theory. There are factors that speak against this narrative, however, the most critical being its timing.

Darwin famously developed his evolutionary theory over the course of several years beginning in 1837 and continuing into the next decade. The best evidence we have of this period comes from his "transmutation" notebooks, which reveal the gradual development of the idea through 1837-8 leading to the Malthusian insight of September 28th, 1838, which presents the first recognizable articulation of natural selection (p. 135).[7] If we follow this timeline, however, Darwin's psychological claims cannot be regarded as product of his evolutionary insights. The "M" notebook, wherein one finds early speculations concerning animal reasoning and morality, was written largely before the storied Malthus reading and long before the theory was worked out. The "Old and Useless" notebook (OUN), which contains some of the most radical reflections on animal reason, will, and consciousness goes back even further. Entries have been traced by their

---

[7] Unless otherwise noted, quotes from the notebooks are from van Wyhe (2002-).

watermarks to as far back as 1837 and a note from Darwin marks the book as containing thoughts from "1837 & earlier," suggesting that some of the underlying ideas could be older still. More tellingly, psychological notes appear early and often in the "transmutation" books themselves. Already in notebook B, the book containing Darwin's first and least worked-out thoughts on speciation, one finds him gathering reports on animals having foresight (p. 165e; p. 255e) and citing his belief that human and animal intellects are not radically different as a major difference between himself and Lamarck (p. 214). Indeed, the inner lives of animals are present from the very first entry, with Darwin commenting that progressive development influences "even mind and instinct" and drawing a parallel between the way "savage" children and wild birds acquire ideas (pp. 3-4). Finally, we have record from May-June 1837—the same sparsely documented time in which Darwin seems to have become an evolutionist (Browne, 1995, pp. 359-60)—of him taking an interest in Crows' wariness of men with guns (an example of animal reason drawn from Erasmus Darwin's *Zoonomia*).[8][9] These documents speak strongly against the idea that the psychological claims of Darwin's later work are mere addenda to the *Origin*. Even the moderate view that they emerge from his "applying" still-developing evolutionary ideas to a new domain (e.g., Gruber and Barrett, 1974, p. 20) seems off, given how old some of the notebook passages are and the fact that Darwin was asking after crows' fear of guns before any of the writings. Physical and psychological, it seems, were together from the start.

Another hypothesis, prominently developed by Richards (1987), is that Darwin's views trace to contemporary sensationalist psychology. On this view, Darwin's willingness to attribute

---

[8] Charles to Caroline Darwin, 19 May-16 June 1837, letter 360

[9] Unless otherwise noted, letters are from the Darwin Correspondence Project, available online at: https://www.darwinproject.ac.uk

complex psychological processes to animals would have been facilitated by his reading of figures like Locke, Hartley, and Buffon. The central vision of sensationalism, or at least the strain that Richards emphasizes, was "that all ideas merely imagined sensations and that rational behavior, of which even animals were capable, derived from habitually associated ideas" (1987, p. 21). Believing that reason consisted in trains of sensations linked by experiential learning, a capacity often granted to animals, the sensationalists offered a psychology naturally sympathetic to reasoning animals. There is an essential truth to this picture. Sensationalism was doubtlessly a major force in British psychology, and there's no denying that many in the tradition, including Hume and Hartley, presented forceful arguments in favor of animal reason (see below). It's also true that the approach is far more congenial to the claims of animal reason than, say, Kantian psychology or the "commonsense" school of Reid and Stewart. As an explanation, however, the appeal to sensationalism alone faces issues similar to those speaking against the evolutionary account.

To start with, there is ample evidence of sensationalist views being held without anything nearly as radical as Darwin's proposals. As one of the most common modes of psychological thinking at the time, sensationalism housed both proponents and detractors of animal reason. Indeed, the critics included some of its most illustrious figures. Locke, widely regarded as the tradition's originator, contended that animals were most likely incapable of entertaining complex ideas or comparing one with another (*Essay Concerning Human Understanding*, II, XI) and denied them "abstract ideas" (*Essay,* II, XI), though later sensationalists would do away with that particular concept. In France, the sensationalist and early evolutionist Buffon was similarly dismissive, insisting that, while animals possessed memory and excellent senses, they could not reason: "In all their operations," he wrote in his *Natural History*, "there is neither reflection nor thought" (p. 69). Lamarck and Condillac expressed similar reservations, the latter arguing that animals' lack of language meant that they could not reflect or control their imaginations (*Essay on the Origin of Human Knowledge*, I, iv, I). The same assumptions, one will note, were made by many of the late 19th century authors who cast doubt on Darwin's psychological work. Lloyd Morgan, for instance, identified strongly with the school (Richards, 1987, ch.8). Although sensationalism was more likely than others to house believers in animal reason, then, it was far from the consensus in that tradition. Even if Darwin were a thoroughgoing sensationalist, we would be left with the question of why he sided with the reason-attributing branch over their reason-denying peers.

Second, there are issues with the timeline. With some exceptions (his grandfather's *Zoonomia* and perhaps Hume; see 1.4.1 below), most of the sensationalist works believed to influence Darwin were consulted only after his views on animal minds were already formed. His notes on Brougham and Hartley appear mostly in the "N" notebook (c. 1838-9), while Leroy's *Lettres Philosophiques* were marked as finished only in 1842 (Books Read, 19v). As we saw in the discussion of his evolutionary ideas, however, there are many comments from before this time that show Darwin holding a firm commitment to mental continuity and the idea that animals reason. The B notebook, with its notes on reasoning farm animals and its self-conscious rejection of Lamarck's exceptionalism, looks to have been completed before any of these readings were undertaken. Likewise, relevant continuity-defending passages from the M, C, and Old and Useless notebooks (see section 1.4.2) all seem older than the readings listed above. That these works influenced Darwin and his mature thought need not be disputed—he clearly saw Brougham, Hartley, and others in that tradition as luminaries—but if we are to understand the genesis of his views, it suggests, we will need to look elsewhere.

Finally, Darwin's conception of reason seems stronger than many sensationalists. If all that was meant by reason were a special kind of association, then it wouldn't take much to credit it to animals. Darwin evidently has more in mind, though. In the 1874 edition of the *Descent*, for instance, he draws a clear distinction between reason and the "mere association of ideas" (1874, p. 74), crediting the latter with explaining a pike's slowly learning to avoid a glass barrier but presenting a macaque's rapid behavioral adaptations as an instance of the former. The same distinction is visible in the notebooks too, with Darwin treating instinct, reason, and "experience, or habit" separately (see M, pp. 76-7) and favorably citing authors critical of the sensationalist tradition. All three metaphysical books evidence a great respect for Sir James Mackintosh, for instance, positively citing work in which Mackintosh blasts Locke's tabula rasa as simplistic and suggests that "many more, as well as greater laws" than those of the sensationalists will prove necessary for an adequate psychology (pp. 201-2). A still more direct example comes from the "M" notebook, where Darwin self-consciously adopts John Abercrombie's expansive definition of reason (p. 114, p. 141), viz. the ability to "compare facts with facts, and events with events; to weigh their relations, bearings, and tendencies; and to assign to each circumstance its proper weight and influence in the conclusions which we are to deduce from them" (1838, p. 137). In the cited work, Abercrombie is—like Darwin—careful to distinguish reason from the central sensationalist building block of association (1834, p. 176, p. 332; it also dismisses the central sensationalist notion of "ideas" as a "fiction" [p. 24]).[10] Darwin was no Kantian, but these

---

[10] Examining Darwin's copy, he seems to take on far more. The human-centered text seems to have been used as a kind of checklist for faculties. Where Abercrombie notes imagination in painters and novelists Darwin notes that "when cat pounces & runs after feather, it knows it is not mouse, but does it not use imagination or picture to itself it is" (p. 163); talk of human abstraction

observations suggest that Darwin's conception of reason was more demanding than the notion adopted by many of his sensationalist influences.

Despite the progress that has been made in accounting for Darwin's views over the past several decades, then, we are still left with a gap. Both his evolutionary thinking and his readings in sensationalist psychology have a role, but they appear insufficient. It is still too easy to see someone accepting both elements while at the same time scoffing at the notion that humans and animals had a shared capacity for reasoning or judgment. Why, one wonders, would Darwin take on the burden of arguing for animal reason—of animals' capacity to abstract, envision future events, and even engage in self-conscious moral evaluation (1871, p. 62-3; 72-3)—if the resources were available for a more modest and far more easily defended position? What is missing, I would like to suggest, isn't a single idea like species transmutation or a philosophical system like sensationalist psychology but something at once larger and more specific. It is, as has been suggested in discussions of his evolutionary theory (Desmond and Moore, 1992; Hale, 2014; Desmond and Moore, 2009), a question related to moral culture. Darwin, as I'll argue, was part of a narrow intellectual and political sub-culture in which his views were not only common but reinforced by a broad set of ethical and religious sentiments. He was raised in a tradition of radical and reform liberalism that had a long history in North and Midlands Britain, the same tradition that would give rise to the first animal welfare activists and the RSPCA. Its sentiments, not quite rising to the level of a philosophical school, represented an eclectic mix of Enlightenment doubts

---

elicits "does not a bird when it builds its nest, use abstraction respecting place, & softness & elasticity of materials, which are not constant in kind, but only in quality" (p. 158); and studies of mental illness feature the note "animals minds are influenced by age, like that of man" (p. 154). This suggests that Darwin was willing to think in terms a bit closer to Scottish faculty psychology.

and virtues, the former developed in opposition to the Oxbridge schoolmen and the latter in line with a growing emphasis on sympathy as the overriding moral guide. The skeptical element would naturally push him away from the kind of god that would make man (and man alone) in his image: like the Unitarians amidst whom he was raised, Darwin was likewise skeptical of a lord that would ordain natural and immutable hierarchies. The saltation story offered by Lyell and others would thus carry little appeal for him. So too, the virtue of "sympathy" or "pity" championed by Voltaire, Godwin, and many of Darwin's friends and relatives would help to motivate the specific levelling project he undertook with respect to the animal kingdom. The uniqueness of human reason was a prop for this that would naturally earn his hostility and, perhaps, act as a spur to action.

1.3 "Thy brother Emmets, and thy sister Worms!"

To understand Darwin's commitment, it's helpful to start with some historical context. Through the seventeenth and early eighteenth centuries—and well into the nineteenth at institutions like Cambridge (Fyfe, 1997)—the received academic view on animals derived from a synthesis of Christian and Aristotelian ideas (Fudge, 2005, ch. 3; Gibson, 2015, ch., 1; Harrison, 1993). All animals, human and non-, were said to possess the sensitive, material souls needed for perception. Yet it was humans alone who had been made in God's image and, consequently, proved capable of reason and subject to the risks and rewards of immortality. While animals felt, physicians and anatomists like Tyson (1699) contended, they could not think as humans do. Lacking reason, moreover, they were—as Aquinas noted—"naturally enslaved and accommodated to the uses of others" (*Summa Theologica*, Pt. II-II, Qu. 64, Art. 1). As a growing body of scholarship has shown, however, the era also saw the emergence of a moral and political sub-culture with a decidedly different take. Aided by broad social trends like the keeping of pets and the circulation of traveler's reports on the philosophies of southeast Asia, animal welfare became an increasingly common concern among members of the educated classes (Thomas, 1983; Shevlelow, 2008; Stuart, 2007). In France, Voltaire would rush to defend the reason of birds and hounds, decrying the "barbarian" anatomists who would "nail [a dog] to a table and dissect him living" ("Beasts"), and La Mettrie suggested that, given the opportunity, an ape might learn language and how to interact in human society, becoming "a little gentleman, with as much matter or muscle as we have, for thinking and profiting by his education" (1748[1912], p. 103). In England, works by the vegetarian Thomas Tryon gained a wide audience for the idea that animals had rights and that there could exist "mutual Contract or Stipulation agreed betwixt Man and [beast]" (1695, 149-50; see Stuart, 2007, ch.5), while the Associationist David Hartley suggested that we could well be "members of the same mystical body" with the same prospects for resurrection (1749, p. 223). (Both men would naturally become favorites of the Romantics, with Coleridge so enamored of Hartley as to name his son after him). In Edinburgh, meanwhile, Hume declared the arguments in favor of animal reason "so obvious, that they never escape the most stupid and ignorant" (bk I, sec. XVI)—a view shared by the Encyclopaedist William Smellie and key members of the nascent anti-cruelty movement hailing from the region, including Lord Erskine, Lord Brougham, and James Mackintosh.

Most importantly for our narrative, however, is a network of thinkers in Birmingham and Derby that included Darwin's grandfather Erasmus. From his home in Lichfield, the elder Darwin maintained ties with a network of scholars, poets, and artists, that extended across Europe to the intellectuals of Paris, London, and Edinburgh. His friends included the anarchist William Godwin, an aging Rousseau, and a young and still very radical Samuel Taylor Coleridge (who believed that

Darwin might possess “a greater range of knowledge than any other man in Europe”; letter to Cottle, Jan. 1796; Coleridge, 1847, p. 64). Closer to home, associates in the Birmingham Lunar Society and Derby Philosophical Society included the Unitarian Joseph Priestly, a materialist ridiculed by Thomas Reid for believing “that plants have some degree of sensation” and animals “differ from us in degree only, and not in kind” (1872, 52), the radical Quaker Robert Bage, whose novels expressed the sentiment that “man is but a genus of the monkey tribe” (1787, 144), and Thomas Day, who died when thrown by a horse whom he’d educated in accordance with nonviolent Rousseauean principles (Basker, 2002, 203; for overviews of the Lunar Society, see Uglow, 2002). Other members of this midlands network included the poet Anna Barbauld (*nee* Aikin), whose poem “the Mouse’s Petition” warned readers to “Beware, lest in the worm you crush/A brother's soul you find/And tremble lest thy luckless hand/Dislodge a kindred mind,”[11] and the extended family of fellow Lunar-man Josiah Wedgwood—Charles’ maternal grandfather and the forbearer of his cousin/wife Emma—who became some of the most active figures in the same Georgian anti-cruelty movement. (During their courtship, Emma was sure to tell her aunt Sismondi that Charles was humane to animals, in addition to being honest and sweet tempered; E. Darwin, 1904, 409).

Out of all his associates, however, Erasmus Darwin was perhaps the most willing to see animals as our moral and mental peers. Our fundamental kinship and similarity is a theme that appears across the doctor’s poetic and philosophical writings. In his monumental *Zoonomia*, a text after which Charles would name the first of his transmutation notebooks, Darwin attributes reason

---

[11] The poem was famously affixed to the cage of one of Priestley’s lab mice. According to an early report from William Turner (1825), the work was moving enough for Priestley to release the mouse but not to halt his experiments generally.

as diverse as wasps and rooks and, like Priestley and his grandson, extends the realm of ideas into the plant world (1794b, I, 105). Echoing Tryon, the text claims that creatures like dogs and cats may enter into contracts and deserve the moral status that capacity affords:

> An ingenious philosopher has lately denied, that animals can enter into contracts, and thinks this an essential difference between them and the human creature:—but does not daily observation convince us, that they form contracts of friendship with each other, and with mankind? When puppies and kittens play together is there not a tacit contract….In the same manner that all contracts are made amongst men, that do not understand each others [sic] arbitrary language. (1794b, I, 169)

His poetic works were, if anything, more explicit. His *Botanic Garden*, for instance, sings the praises of a vegetable and grain-based diet, describing "the destruction of other living animals" as the lowest mode of attaining sustenance and celebrating the millstone as providing humanity with a "[f]east without blood!" (1794a, Canto I, line 278). Likewise, his posthumously published *Temple of Nature* presents a vision of the animal world that is richly interconnected and psychologically complex. In it, he celebrates animals:

> Wise to the present, nor to the future blind,
> They link the reasoning reptile to mankind!
> Stoop, selfish Pride! survey thy kindred forms,
> Thy brother Emmets [ants], and thy sister Worms! (Canto III, lines 432-4).

Though humans, ants, and worms are far from identical in terms of their capacity for understanding, all three possess its fundamental components. All experienced sensations, formed associations or trains of ideas, and expressed their will through volitional action; all saw into the future, though perhaps only murkily. At points, Darwin even speaks of a quasi-spiritual "transmigrating ens" (*Botanic Garden,* Canto I, line 574) that, while material in nature, suffuses all living things and unites them within a greater "Ens Entium" (*Temple of Nature,* Canto IV, note to 453). He sometimes identifies this cosmic force as 'IMMORTAL LOVE,' crediting it with having "vivified all things, producing life and joy" (*Botanic Garden,* Canto I, line 416).

It would be difficult to find a single philosophical current that fit each figure. They differed sharply in terms of epistemology and metaphysics: some were materialists and others mystics; some believed that all ideas came from experience, others felt that divine revelation was available to all living creatures. Many, including Darwin himself, gladly switched between vocabularies as it was convenient. Ultimately, though, these differences didn't matter because the network wasn't defined by abstract philosophy but by moral vision and culture. Virtually all those referenced above can be classed as nonconformist in their religion and reformers or radicals in their politics.[12] Priestley was a Francophile who denied the complete veracity of the Bible; Mackintosh was a Whig MP active not only in anti-cruelty legislation but in the abolition movement, electoral reform, and (unsuccessfully) the end of capital punishment—and these were the moderates. The more radical wing included the anarchist Godwin, feminist Mary Wollstonecraft, and utopians such as Samuel Taylor Coleridge and Robert Southey (whose "Pantisocracy," or rule by all, explicitly included a place for animals; see Coleridge's "To a Young Ass").[13] Where Erasmus Darwin's sympathies lay exactly is difficult to know, but he was an enthusiastic reader of Godwin, an ardent abolitionist, a believer in women's education, and openly supportive of both the

---

[12] Hume is an exception, though few conservatives of the day would've taken him as a flagbearer.

[13] Both Coleridge and Southey's politics would change as they grew older, but the latter seems to have held out hope for a place where "yet untaught to bend his massive knee/Wisest of brutes, the elephant runs free" ("The Poets Pilgramage to Waterloo," II, IV, 46).

American and French Revolutions (Desmond and Moore, 2009, ch.1; Hale, 2014, ch. 1). The last cause, it's worth adding, was sufficiently controversial at the time to trigger riots in which Joseph Priestley and other Dissenters' homes were burned to the ground.

The belief that animals thought as humans do was not exclusive to such figures—the Tory Alexander Pope adopted similar stances on occasion—but there are several reasons why the belief was associated with political currents. The first is broadly religious. At the time, reason-based human exceptionalism was a major part of conservative Anglican theology. Reason was tied not only to life after death and moral culpability but the divine organization of authority in society (Lovejoy, 1936; Candland, 1995). As rational beings *par excellence*, adult men stood higher than women and children and whites stood higher than colonized and enslaved peoples (see Strong, 2007, ch. 2 for a discussion of Anglican Imperialism). Animals, of course, stood lower still, though they were frequently invoked for the purposes of dehumanization. For those wishing to uphold these structures, the role of reason in ordering the "great chain of being" would prove difficult to work around (as evidenced even in the Times' review of the *Descent*). Pope, it will be noted, avoids the conflict only by denying animals immortality and insisting upon an "insuperable line" between us (*Essay on Man*, Epistle 1, line 228). For the liberal Unitarians, Quakers, and Deists, however, contiguity would be the more appealing relation between the various groups. Though by no means egalitarian, the Dissenters' assumptions inclined them toward a more robust rhetoric of brother and sisterhood, as seen in debates over slavery, the education of women, and the place of women and non-whites in religious observances (e.g., being allowed to speak in services). Having taken this stance in other contexts, it would be more difficult for them to switch so abruptly to the rhetoric of reason when discussing animals. Talk of "insuperable" lines—wherever it appeared—couldn't help but call forth suspicion.

The second might be called moral or humanitarian, though it would be much intermixed with the theological dimension sketched above. Frequently, defenders of animal reason were vocal advocates of what later critics called the "new morality" (figure 1), an ethical impulse setting a' universal love" (*Temple of Nature*, Canto III, VIII)[14] higher than the virtues of national or religious loyalty. Love of king fell to visions of ideal republics or radical experiments like Coleridge and Southey's Pantisocracy;[15] *Amor Patriae* ceded ground to an internationalism and trade; reverence for the church and its prerogatives gave way to "liberty of conscience" and calls for the abolition of state religion. In some cases, the bonds of family itself dissolved before the impartial judge, with figures like William Godwin claiming that, should one have to choose whom to save from a fire, the life of a philanthropist stranger held greater claim than that of one's wife, mother, or benefactor (1793, p. 82-83). At the time, no legal protections existed for non-human animals, and standards for what constituted morally acceptable treatment differed markedly from what's now taken for granted. It wasn't uncommon for young boys to shoot cats as a pastime or for horses to have an eye knocked out or a skull fractured by a frustrated carriage driver (Shevelow, 2009, p. 133). This was more than could be overlooked by someone founding their morality on fellow-feeling, and while there are obviously ways to oppose such practices without going so far as to say that animals think and reason, the attribution is a natural and polemically useful response. That

---

[14] The context of the "Progress of Mind," an episode in which "Seraph, Sympathy, from Heaven descends" and in doing so "Unbars the prison, liberates the slave" and "charms the world with universal love" (Canto III, VIII).

[15] Those in Britain who remained loyal to the monarchy, such as Hume, did so not out of a respect for the institution so much as a wariness of the risks posed by revolution. "Hume," the conservative Samuel Johnson famously remarked "is a Tory by chance…not on principle of duty" (Boswell, 1786, p. 309).

they should feel was enough for many moderate reformers, but such arguments could only go so far. Many at the time would, like Doctor Darwin's "ingenious philosopher," ground their moral considerations solely on the possession of reason or contract. Others would downplay animals' suffering, as it would not be accompanied by dread or an awareness of their situation. Pope's *Essay on Man*, for instance, praised the Lamb's "blindness to the future" as a mercy granted to beings on whom humans fed (Ep. I, line 85); "Had he thy reason," the poet asked, "would he skip and play" (line 82). Whether they were "wise to the present" or "to the future blind," Darwin perceived, mattered a great deal even when the discussion was between animal sympathizers. In terms of strategy, then, the ideal position for one sympathetic to the animals would be to attack on all fronts. In practice, advocates from Bayle and Voltaire to the elder Darwin did just this. Pain is accompanied by a measure of understanding and foresight; Animals deserve our respect as fellow thinkers as well as our pity as fellow feelers.

Figure 1. James Gillray. "New morality; or the promis'd installment of the high-priest of the Theophilanthropes, with the homage of Leviathan and his suite." © National Portrait Gallery, London.

In the background, behind the "Cornucopia of Ignorance" and beside a young Coleridge (pictured as an ass, in reference to the poet's "To a Young Ass"), one can see Erasmus Darwin carrying a basket of liberty-cap wearing vegetation labelled "Zoonomia; or Jacobin Plants." Other notables include Joseph Priestley (ahead of the green "Leviathan"), Lord Erskine (in the water near Leviathan's tail), Robert Southey (the ass below Coleridge), and William Godwin (also an ass). The cartoon additionally references Voltaire, in the adjoining poem, and Sir James Mackintosh, via a paper labelled "Concel McFungus Speech." Note that "Sensibility" in the upper righthand corner is presented as mourning a dead bird while standing on the head of a decapitated king.

1.4 Enter Darwin

At the time of the younger Darwin's writing, it was common to see connections drawn between his work and the sort of figure discussed above. A review in *Blackwood's,* for instance, labelled the Descent "A Continuation of an Old Song," claiming in verse that his grandfather's *Loves of the Plants* "must have softened his heart, when his bosom was young" and that the *Temple of Nature* "has prompted his tongue" (1871, p. 518). Other ties were made to Hume, whose school took so seriously "the old stories about the reasoning of animals" (Müller, 1887, p. 119), and Smellie, alongside whom Darwin made the "foolish" case "that the instinctive operations of animals result from intelligence and reason" (Cooke, 1877, p. 185). Some of the less sympathetic reviewers drew connections to the Whig Deist Lord Monboddo, ridiculed for believing that humans once had tails and, like La Mettrie, that apes might be taught language and culture (playing the harp, for instance; see Monboddo, 1774, p. 276). Darwin was generally careful to avoid such connections, and understandably so. Beyond undercutting the originality of his contributions, a nod to these figures could easily render his work less scholarly in the eyes of his peers. Darwin didn't want his ideas dismissed as literary the way Robert Chambers' anonymously published *Vestiges of the Natural History of Creation* had been (Secord, 2001; 2015), and too strong an association with the author of the *Botanic Garden* could very easily render this the case.

Nevertheless, I claim, these traditions did influence Darwin. There are two lines of evidence that lead me to this conclusion. The first comes from Darwin's upbringing and political identity. As a child, Darwin would have witnessed the reserved freethinking of his father and the social activism of his sisters and cousins, and when it comes to animals, there is little debating that Darwin's heart was softened. "Regard to suffering in both man and beast," his son Francis recalled, "was indeed one of the strongest feelings in his nature" (Darwin, 1887, p. 200). As with the previous generation and for essentially the same reasons, I suggest, these facts would predispose Darwin to view animals as beings with complex inner lives. The second line of evidence stems from the notebooks. Darwin was kind enough to leave one of the best paper trails in the history of science, documenting his readings and thoughts through much of his adult life—including, as we saw earlier, the critical years in which he formulated his central ideas. Looking to these notes, we find an intimate familiarity with his grandfather's work, expressions of moral outrage, and a keen interest in the tradition discussed in the previous section. This background, I claim, informs his selection and use of sources in several obvious ways, revealing partiality and purpose.

1.4.1 Early Life

Given that Darwin's continuity-suggestive remarks do begin so early, it makes sense to look first to his upbringing. Though more moderate than his Grandfather's publications might imply, Darwin's education was doubtlessly unconventional. Mindful of the previous century's riots, his parents thought it best to raise their children as Anglicans. At home, however, Charles had every opportunity to develop sympathies of a very different sort. His father Robert was a Unitarian and, according to Darwin, a "freethinker" in the mold of the previous generation,[16] while his mother Susannah encouraged liberal tendencies by bringing him to services by the Unitarian minister George Case—going so far as to hire the reverend as a tutor for Charles (Darwin, 1958, p. 22; Browne, 1996, p. 12). His "radical sisters"[17] were all socially active Whigs, and after the

---

[16] Darwin to Galton, May 1873, letter 8924

[17] Darwin to Fitzroy, October 1836, letter 310

untimely death of their mother, were largely responsible for his education. By all accounts, this included a hefty moral component. On Darwin's telling, his sister Caroline was so "zealous in trying to improve [him]" that he scarce entered a room without calling to mind a wrongdoing he might be blamed for (1958, p. 24). In terms of what he studied, there is some indication that the *Blackwood's* claim was correct about the softening of his heart. The young Darwin was exposed to his grandfather's poetic work, for instance, complete with themes of animal kinship and divine sympathy (Browne, 1995, p. 15), and by his own recollection, he would return to the works of animal-lovers like Shelley, Byron, and Coleridge many times (1958, p. 43, p. 138). It's likely that he would have been familiar with Barbauld also, since her writings had been a staple of the Wedgwood home (Uglow, 2005, p. 169) and were part of Erasmus' curriculum for the moral education of women like Charles' sisters (see Darwin, 1797, p. 126). If this latter work is any guide, Darwin could have been exposed to a long line of animal sympathizers, including Linnaeus, Ovid, Pope, John Aikin, and others. As he grew older, these tendencies would be reinforced through personal readings as well as the negative example provided by his much-despised stay at the more traditional Shrewsbury school of Bishop Samuel Butler (1958, p. 27-8).

Outside the home, Darwin spent a good deal of time at the estate of his uncle Josiah Wedgwood II, one of the premier philanthropic households of the day. Particularly active were the women of the family, including future-wife Emma, who organized outreach to infirmaries and asylums, hosted temperance teas, and helped establish Birmingham's local anti-slavery society (Crisafulli, 2010; Desmond and Moore, 2009, p. 13). The cause of animal welfare was naturally another going concern for the Wedgwoods. It was in their household that Darwin learned, when fishing, to kill worms with salt water rather than spit them live on a hook (1958, p. 27) and, more substantively, where he witnessed early animal welfare activism. The Wedgwood girls' maternal aunt and regular visitor Catherine Mackintosh (wife of James) was a founding member of the SPCA and kept the family abreast of developments in that arena. In 1826, one letter reports, "Kitty" was busy "getting a man convicted of cruelty to his ass" and imprisoned on grounds of newly passed legislation protecting farm animals.[18] In later years, the Wedgwood clan would become active in the anti-vivisection movement[19] and even make it into British case law in Allen v Wedgwood, a ruling that held animal welfare to be a valid charitable trust in the execution of an individual's will.

As a young adult, it seems, Darwin's affiliation with such causes only grew. We know from his letters that he was also introduced to more than one Whig statesman, including James Mackintosh.[20] The statesman was busily working on behalf of animal welfare at the time, including the aforementioned case of the ill-treated ass, and as Darwin describes their first meeting in the *Autobiography*, "I listened with much interest to everything which he said, for I was as ignorant as a pig about his subjects of history, politicks and moral philosophy" (1958, p. 55). Should the matter of animal minds have been broached at this or later meetings, as seems likely given Darwin's interests and Mackintosh's activities at the time, we also know what the statesman's take would have been: for Mackintosh rationality could be found in creatures "extending from Newton to the elephant" (1835, p. 36). It was around this same time that he and elder brother Erasmus

---

[18] Mrs Josiah Wedgwood to her husband, 26 November, 1826 in Litchfield (1904, pp. 246-7)

[19] Darwin's position on the matter was complicated. He thought it necessary for the greater good but was personally horrified at research undertaken for "damnable and detestable curiosity" (see Darwin, 1887, p. 200). Similarly conflicted feelings attend his thoughts on hunting.

[20] See Catherine and Caroline Darwin to Charles, 11 April, 1826, letter 31 and Darwin to W.D. Fox, 3 January, 1830, letter 75.

began reading, as the latter put it, “like horses.”[21] They borrowed more books from the university stacks than any other students (St James’s Gazette, 1888) and, judging from surviving notes and letters, obtained still more from lending libraries. Erasmus’ record indicates works by Hume and Voltaire, likely shared with Charles, while Darwin’s own reading notes indicate a first reading of *Zoonomia*, which he “admired greatly” at the time (1958, p. 49). Looking to his school peers, chances are that he would not have been alone in the admiration. Edinburgh’s medical men were notoriously radical (see Desmond, 1989), including as many materialists and freethinkers as anywhere in the British Isles. Pamphlets, journals, and extramural lectures discussing materialism and evolutionary ideas could all be found quite readily. Darwin’s mentor Robert Grant, for instance, was a Lamarckian whose lectures celebrated nature in terms not unlike old Erasmus’ poetry and whose comparative anatomy simply assumed animal intelligence lay behind things like vocal communication (p. 523). Darwin himself participated in several of the city’s student groups, including the Plinian Society, where he could here presentations like fellow-student William Browne’s defense of materialism and William Rathbone Greg’s case that “the lower animals possess every faculty & propensity of the human mind” (Browne, 1995, p. 78).

Even before his famed voyage, then, Darwin would have had ample exposure to the idea of mental continuity and the sentiments historically associated with it. Those who have the most claim to being his early role models—his father and grandfather, sisters and cousins, Mackintosh and Grant—were all members of a moral culture that, while far from homogenous, was firmly committed to the idea of animal reason and skeptical of any claims of human exceptionalism. Where other naturalists would have been educated to see humanity as uniquely crafted in God’s

---

[21] Erasmus Darwin to Charles Darwin, 24 February, 1825, letter 13

image, Darwin would have been taught to see the mouse, ass, and worm as his brothers; while they were taking in orthodox theology and one or another form of divine creation, Darwin was hearing about materialism and the transmigrating Ens. Considering his subsequent work, it's hard to think this assumed kinship didn't play a part. It's a small step from moral brotherhood to mental or material kinship—one that a great many of his peers were willing to make—and looking to the crucial period following his return from abroad, we find ample reason to see it as just this.

1.4.2 "One thinking sensible principle"

Correspondences aside, the voyage years placed Darwin at something of a distance from the liberal surrounds of his upbringing. As his old congenial acquaintances relished in reminding him, Darwin was in conservative company.[22] Upon his return, however, matters returned to their usual state. As he had in the early years of medical school, Darwin also spent a great deal of time reading and socializing with naturalists, including animal sympathizers at the Zoological Gardens. It's also the time in which he begins putting his thoughts and academic researches to paper. The seeds of the *Origin*, the *Descent*, and the *Expression of Emotions* were all sewn over the course of a few years, and it's from the records of this time that the moral-theological element gains its strongest support.

---

[22] One letter jokes that his being "amongst a Tory Crew" might allow him to make a good zoological specimen—bonus if he could catch one with elongated ears or "Monboddo's Tail" (a reference to both early evolutionist ideas and the philosopher's belief that humans once had them; Hooker to Darwin, April 15, 1832, letter 165).

To start with, one finds morally loaded terminology interspersed throughout the books. Darwin repeatedly likens the position of animals to slavery (B, pp. 231-2; C, p. 154), juxtaposing this with the familiar image of animals as brothers, sisters, and friends. Animals are “our fellow breth[re]n in pain, disease, death & suffering, & famine” as well as “companions in our amusements” (B, p. 232); they show “affections, imitation, fear of death, pain, [and] sorrow for the dead” and as such are deserving of respect (C, p. 231). In notebook C, he expresses the thought that:

> Animals have voice so has man. Not <u>saltus</u> but <u>hiatus</u> hence if sickness death, unequal life — stimulated by same passions, brought into the world same way, animals expression of countenance. They may convey much thus. Man has expression. — animals signals, (rabbit stamping ground), man signals. — animals understand the language. They know the cry of pain as well as we. (C, p. 154)

For Darwin, it’s only “our arrogance” (C, p. 154) and delusions of a “godlike” origin (C, p. 155; see also p. 196) that prevent us from seeing these commonalities.

Such remarks are telling for at least two reasons. First, the language itself is emotionally charged and quite similar to that seen in the previous section. His comparison of animal usage to slavery and insistence on the arrogance of those who deny our kinship express clear moral indignation and could easily have come from Coleridge, Barbauld, or his grandfather. This suggests that the culture they represent and which Darwin was a part of throughout upbringing were still active at the time, pushing toward mental continuity as one aspect of a larger moral vision (a vision that seems to have remained well into Darwin’s old age; see, F. Darwin, 1887, p. 200). Second, the remarks are found in the transmutation notebooks, suggesting a clear tie to his natural philosophy. One can imagine a scenario in which his moral/political side was largely separate from his interests as a naturalist; people often have varied interests, and these don't always interact. Were this the case for Darwin, however, there would be little reason for such remarks to take up space in the B and C notebooks. The transmutation books are, for all the varied thoughts expressed within them, topical in nature. They don't contain the sorts of aphorisms and free thoughts one finds in many commonplace books (e.g., Lichtenberg's), but instead appear focused on developing his ideas on natural history. If these morally loaded entries linking fear of death and sorrow for the dead to worthiness of respect are found there, then, it's presumably because they were, in fact, connected in his mind.

Turning from the moral to the spiritual, a similar set of observations holds, with Darwin's notes showing more than a passing resemblance to the panpsychist and pantheist undercurrents of his predecessors. Examining the M, N, and OUN notebooks, for instance, one finds that Darwin, like Priestley, his grandfather, and Shelley, suspects that plants have a psychology of sorts. He wonders whether they may have a notion of cause and effect (N, p. 13), speculates on whether they have free will (M, p. 72; OUN, p. 35), theorizes about the basis of their "memory" (N, p. 50), and writes of "my theory of sensitive plants" (N, p. 50; Cf. OUN, p. 9). The mystical element can be seen in his views on animals too. On a few occasions he speculates that memory may be passed down from generation to generation in the manner of Erasmus' transmigrating Ens, allowing for a kind of afterlife and helping to explain mental similarities across generations. In OUN, for example, he asks: "is it more wonderful that memory should be transmitted [intergenerationally]…than from hour to hour in man," crossing out the last word to write the more inclusive "individual" (p. 37). On the same page, he allows that the materialism he envisions may be theistic and that higher animals—those possessing memories—may yet attain a "reward in good life" through their descendants (p. 37; Cf. M, p. 123). In notebook C, he further links this "memory in offspring" to the experience of recalling something after a long period of time and, though he thinks so specific a transmitted memory is unlikely, to the experience of déjà vu (p. 172). Like his grandfather before him, then, the young Darwin seems to be materializing the community and transmigration of souls. Each generation passes some of its mind on to its progeny, linking each creature in a brotherhood of mind. At moments, there is even something like the Ens Entium: "there is," Darwin writes in notebook C, "one living spirit prevalent over this world…which assumes a multitude of forms each having acting principle according to subordinate laws," as well as "one thinking sensible principle (intimately allied to one kind of organic matter…which is modified into endless forms bearing a close relation in degree & kind to the endless forms of the living beings" (C, p. 210e). Again, human and animal minds were not only similar but, potentially, united in a greater nexus of sensibility and thought. In keeping with the tradition of his grandfather and the Romantics, moreover, he seems to have extended this to the plant world. The "acting principle" that underlies the "community of mind" is, the passage notes, something that unites man with man, polyp to polyp, and bud to bud. (It also contains cryptic but suggestive remarks about animal magnetism and sleepwalking, favorite topics among Romantic-era pantheists and, in the case of sleepwalking, Darwin's M notebook).

As with the moral remarks, this suggests a strong connection between the ideological undercurrents discussed earlier and Darwin's views on mental continuity. Few people took seriously the idea of minded plants or a vast spirit uniting the living world; it was the sort of thing that got one ridiculed, in fact. Those who did—Erasmus Darwin, Percy Shelley, Joseph Priestley, the Quaker naturalist William Bartram (also read by Darwin; Books to be Read, p. 5v)—were overwhelmingly members of the tradition sketched earlier. That Darwin echoes these views in the notebooks serves to confirm an identification with that group and a cultural backing for his commitments in animal psychology. The more spiritually intoned elements would give way as he became more agnostic—the emphasis on the OUN being "useless" suggests that he came to have a low opinion of its contents—but they speak to the origin of his later, public commitments as a transitional fossil might speak to the history of a living creature. As his lifelong suspicions about sensitivity and learning in plants (see Darwin, 1880, ch.12) suggests, moreover, the impression stuck in many if not all respects of his psychological thinking.

A third line of evidence comes from the bent of his readings. Examining the works noted and referenced, one finds a noticeable skew toward the types of figure discussed in section 3. Listing only those who advocated or adopted ethical vegetarianism—a distinct minority even among those concerned with animal welfare—one finds Percy Shelley, William Godwin, Jean Jacques Rousseau, William Gardiner, John Hawkesworth, David Hartley, Lord Byron, and, of course, his grandfather. Expanding the list to include those expressing direct concern for animal welfare, it takes on Hume, Smellie, Mackintosh, Montaigne, Brougham, Soame Jenyns, William Youatt, John Lawrence, and many more. These readings often bear directly on the issue of animal minds, and in several cases, one finds that works sympathetic to mental continuity have been taken up or underscored on recommendation of other continuity-defending authors. J.S. Bushnan's *Philosophy of Instinct and Reason* was picked out from its having been dedicated to Brougham (Books to Read, p. 11v); Hartley's *Observations on Man* was "recommended by Mackintosh & Priestly" (p. 13r); Pierquin's treatise on the madness of animals features the note "said to be good by Dr. L. Lindsay" (p. 1r), an early supporter of Darwin and author of the *Mind in the Lower Animals in Health and Disease*; and Delabere Blaine, a veterinarian of "early and enthusiastic attachment to the brute creation" (1824, p. iii), is selected on the basis of praise from fellow advocate William Youatt (p. 16v). In one instance, Youatt's *Obligation and Extent of Humanity to*

Brutes, a straightforward work of advocacy was consulted because it "discusses their minds" (p. 16v).

Darwin's reading of these authors was, as previously noted for the sensationalists among them, too late to be regarded as the *cause* of his views. Yet they provide important clues about his method and frame of mind in approaching the topic. First, they reveal a man with a specific political identity. Darwin's readings, to say nothing of the moral condemnations and metaphysical speculations peppering the notes, are in line with what one would expect from a radical Whig of the time. Many of Gillray's "New Morality" figures, including Godwin, Coleridge, Wollstonecraft, Mackintosh, and Priestley make appearances; had the cartoon been drawn a few decades later, it doubtlessly would've had more. Given that there is such a bent in favor of continuity and that so many of the texts consulted on animal minds were taken up on recommendation of animal-sympathizing authors, moreover, we may venture that Darwin approached the matter with intent. He sought out works that would support his morally preferred position. In fact, when one looks to the notes eventually taken from these works, one finds that his manner of recording arguments and illustrations resembles nothing so much as the stockpiling of ammunition. In the same way that he takes down a report of pigs unhitching the latch from a gate (B, p. 255e) or a dog toying with its caretakers (B, p. 165e), he notes that Brougham has an argument in favor of abstraction (N, p. 62) and that Algernon Wells believes that animals' imitation and educability imply a degree of reason (N, 69-70). There are authors that he engages with more systematically, of course (e.g., Mackintosh in the OUN), but more often (as with Hume [N, p. 101], Aimee Martin [OUN, p. 8], and Brougham), they seem to have been used instrumentally.

Finally, we have his engagement with more orthodox authors, which bear the same marks of intent and antecedent commitment. For example, Darwin seems genuinely surprised that the entomologist George Newport would dispute his grandfather's assertion that wasps clip the wings of flies out of foresight. "Good heavens," Darwin writes, "is it disputed that a wasp has this much intellect" (M, p. 63e). Likewise, upon reading Lyell's (1830, p. 42) claim in the *Principles* that elephants had been endowed by the creator with a special "sagacity" for taking instruction, Darwin seems not to comprehend why their ability to enter into "contracts" wasn't obvious—even his cousin's dog understands the basics of quid pro quo (M, p. 28; Cf. *Zoonomia I*, p. 165 quoted above). At times, this disbelief transitions into outright vexation, as when Darwin groups the instinct-mongers with deniers of fossil evidence:

> But it will be said there are latent ins[tin]cts, — as crows against man with gun, & Bustards, &c., &c.!!!
>
> An American & African form of plant being found in Tristan D'Acunha may be said to deceive man. as likely as fossils in old rocks for same purpose!! (B, p. 218)

Those who deny the intellect involved in the crow's caution, it seems, are similar in Darwin's eyes to those who think geologists are being tricked by demons. The reactions are revealing both for the degree of surprise they document and their emotional valence. That wasps had foresight, crows anticipated the danger of guns, and all manner of animals entered into contracts were not only positions that Darwin held but positions he felt were obvious—positions whose denial could come only from dogmatism. Put another way, Darwin seems to be coming to the discussion with a set of assumptions common among people like Tryon and his grandfather but well outside the norm for naturalists of his day (represented by the likes of Newport and Kirby).

1.5 Making Sense of Darwin's Psychology

Taken as a whole, the evidence reviewed in the previous sections paints a distinct picture of Darwin and the development of his ideas. Religiously, Darwin long held a distaste for the "revengeful tyrant" God he associated with his peers' Christianity (1958, p. 85). In its stead, at least until his eventual lapse of faith, stood a theological outlook closer to that endorsed by the Deists, Unitarians, and liberal Quakers of his grandfather's circle: the god of Darwin's notebooks was not an interventionist tinkering with individual souls—imbuing humanity with a reason exempt from the natural order—but a being that had imbued the world with a common vital-material spirit (OUN, p. 6v; C, pp. 210e-211). The human mind, however impressive it might be, was ultimately made of the same stuff that fashioned plants and animals. Politically, he was steeped in a culture that, compared to most of his peers, was skeptical of received hierarchies and, in his family's case, committed to animal welfare. For one accepting such moral and religious ideals, the thought of human and animal minds being alien would hardly have occurred. As a materialist, there would be no soul "superadded" that might separate us, and a law giving God would not deign to make such exceptions in the first place. To claim that man had been made in God's image would be little more than a pretext for the mistreatment of one's fellow creatures. What's more, these beliefs were not simply private inclinations but commitments visible in his research work. They are to be found explicitly in the moral and metaphysical pronouncements of the B, C, M, and N notebooks, as well as implicitly in the naturalist's choice of sources and reactions to countervailing theories. Though his cultivated image was that of a cool observer, his manuscripts reveal a striking measure of passion.

From the materials we have, this set of commitments seems to have remained in force well into his later years. On more than one occasion, his son Francis would recall, the typically reserved

Darwin would exhaust himself by “violently remonstrating with” someone over the ill-treatment of a horse (1887, p. 200). In the drafting of the *Descent*, *Expression of Emotions*, and the instinct chapter of the *Origin*, moreover, we know that Darwin made heavy use of his notes. New observations would be added to the supporting points found in the notebooks, but in core themes, they are much the same. They share between them the same vision of mental continuity; the same distinctions between habit, instinct, and reason; the same intertwining of these mental powers in conscience and morality; and even slivers of the sensitive plants and moral indignation prominent in the notebooks (e.g., his decision to include the dog who licked the vivisector’s hand in the Descent). In tracing Darwin’s family history, upbringing, and process of research, then, we are brought back to the distinctive set of claims with which we started.

In admitting the moral and cultural dimension of these works, moreover, we find natural responses to those questions left open by accounts centered on his evolutionary views and relation to sensationalism. The issue of timing is easier to explain. As was noted in section 2, the development of his evolutionary thought and his sustained engagement with sensationalist psychology in the late 1830s appear either contemporaneously or after the earliest signs of sympathy for mental continuity. He would have been familiar with these ideas, of course—the environment in Edinburgh and his private readings of *Zoonomia* ensure this much—but neither seems to have engendered any firm commitment at the time. Darwin wasn’t an evolutionist until his return from the voyage, and it’s doubtful he ever fully identified with the sensationalist tradition, as his later work shows a great sympathy for “knowledge without experience” (OUN, p. 33), and his early influences would have included figures like Mackintosh and the rabidly nativist Edinburgh medical men. When it comes to his moral sentiments and religious identity, however, we can say with confidence that they were being pushed toward continuity from the outset. The idea of animals' inner lives being like our own would have been second nature to him even as a student. It's also consistent in an interesting way with his endorsement of instinctive morality. Though orthodox sensationalism has little to recommend them, inborn morality would be a natural byproduct of the idea that all living beings were united by a common thread and that something of our minds is passed down to our progeny.[23] (This went for vices too: as Darwin wrote in the M notebook, "[o]ur descent…is the origin of our evil passions!! — The Devil under form of Baboon is our grandfather!" [p. 123]).

Finally, it helps set Darwin and his sympathizers apart from skeptical contemporaries. We know that human-excepting naturalists like Mivart, though willing to adopt some version of evolutionary theory, were raised in far more traditional homes and were committed to a far more traditional conception of God and the cosmos than Darwin; the same can be said of Lyell (Bartholomew, 1973), whose stance on human exceptionalism was far more tempered than most of his time but nevertheless seen by Darwin as "too orthodox, except for the beneficed clergy."[24] When one looks to those who had similar upbringings and educations, however, the opposite trend holds. Of Darwin's prominent sympathizers and the founding figures of comparative psychology, a surprising number were animal advocates, Edinburgh medical men, or naturalists of liberal dissident leanings. The Edinburgh physician William Lauder Lindsay and famed agnostic T.H. Huxley, for instance, both wrote prominent works defending the mental continuity thesis around the same time as Darwin, while veterinarians like Wesley Mills counted among the most prominent

---

[23] This is not to say that one *couldn't* be a sensationalist and adopt this theological/spiritual view—Erasmus obviously held both—but they are in tension insofar as sensationalism opposes unlearned ideas.

[24] Darwin to Lyell, 9 October, 1866, letter 5236

comparative psychologists in the immediate post-Darwin era. Indeed, when the Society of Comparative Psychology published its first opening address, it was done with support from the SPCA (McEachran, 1888). What holds of Darwin, this suggests, may hold of the founding generation of comparative psychology more generally. Determining whether this is so will ultimately require further research. Given the connections and ideas reviewed here, however, it seems safe to say that the intellectual ancestry of comparative psychology is more expansive and colorful than widely recognized. In thinking about its origins, we need to consider a wider range of figures and attend more closely to the cultural factors that would render the idea of a cross-species mental comparison appealing. Should we continue to see Charles Darwin as the “father” of comparative psychology, it will mean recognizing a lineage and way of doing psychology that is political, spiritual, and, in some ways, poetic.

2.0 Second Paradigm: How Comparative Psychology Lost its Soul—Psychical Research and the New Science of Animal Behavior

From Dewey and McDougal, I turned to my friend Cattell, asking him if his fellow psychologists had really lost the soul. He replied: "I can talk more intelligently about any other subject than the soul. It is well known that psychology lost its soul long ago and is said now to be losing its mind." Anonymous, James McKeen Cattell Papers

2.1 Introduction: A Tone Shift

However we choose to think of the Darwinian approach to comparative psychology, of course, it didn't last. Indeed, as it stood in 1920, the discipline wouldn't have been recognizable to its founders. Methodologically, it had transitioned in fewer than thirty years from an area dominated by field observations, case studies, and at-home experiments to one consisting of lab work and specialized instrumentation. Theoretically, it had gone from a field in which one could discuss "reason," "play," and even "criminality" in animals (Lacassagne, 1882; Lindsay, 1880; Groos, 1898) to one where such notions were looked upon with great skepticism. Culturally, it had gone from a field where the label "anthropomorphic" served as a simple cautionary term to one where it represented an "epithet which you could hurl at an antagonist…when you wished to damn him in genteel terms" or imply that he "was obtuse in intellect, a logical pervert, and an insipid sentimentalist" (Carr, 1927, p. 88). Indeed, one could say that the very object of study had changed. "Of recent years", E.B. Holt observed at the time, "many of the workers in animal psychology have been coming to call this the science of behavior, and have been dwelling less and less on the subject of animal ‘consciousness’” (1915, p. 366).

Within the field itself, this transition is often presented with an air of inevitability. The early comparative psychologists had fallen prey to anthropomorphism, and later authors offered a corrective (albeit one with certain excesses; Galef, 1998). In older histories, the move is presented as little more than the introduction of sound experimental principles (Boring, 1950). Why it began when it did is rarely addressed, however, and when one examines the writings of earlier authors, I argue, one finds that their credulity and methodological naivety has been much exaggerated. Many naturalists working the 1870s-90s were wary of anthropomorphism, encouraging critical distance, experimentation, and attention to animals’ underlying physiology as means of checking the impulse. Were the issue primarily about avoiding interpretive excess or incorporating experimentation, moreover, later psychologists could have found room under the auspices of traditional comparative psychology. There would be little reason for new journals with new vocabularies to be formed, for the bounds of respectability to be drastically redrawn, or for the field’s history to be divided into times “before” and “after” (Waters, 1934).

Looking to the historical context, there are a variety of factors that seem relevant. The professionalization of psychology and the push for “practical” education in the American university system, for instance, have long been recognized as contributing to the more formal, lab-based approach that emerged during the period (O’Donnell, 1985; Wilson, 2002). Examining the extreme rhetoric and measures that attended the shift, however, it seems that an additional factor was likely involved. Focusing on the programme developed by hardline experimentalists like Edward Thorndike and J.B. Watson between 1898 and 1920, I argue that a major catalyst for the transition may be found in the anxieties surrounding the field’s “occult doubles” (Leheay and

Leheay, 1983): psychology reacted so strongly against naturalistic observation, at-home experiment, and anecdote in part because these methods—and much of comparative psychology itself—were associated with spiritualism and psychical research. As a great deal of historical work has gone to show,[25] psychology was, at the time, embroiled in a struggle over the merit of research on telepathy, spirit-communication, and other "supernormal" phenomena. For advocates like William James, the question was one of open-mindedness—a good empiricist would hear the case out, following the evidence wherever it led. For opponents like James Cattell, they represented a serious threat. To openly discuss such matters risked lending credibility to popular irrationalism and the hucksters preying on it. The power this debate exercised over the field is difficult to overstate, and animal psychology presents no exception.

I begin with a discussion of the textbook narrative, arguing for a rethinking of early comparative psychology's methods and the precise nature of the change it underwent. Common assumptions about early psychology's lack of experiment and reliance on anecdote are highly exaggerated, and even if they weren't, it would be difficult to account for the emotional intensity of the transition solely in terms of its participants' methodological preferences. Something else is needed. Turning to the historical literature, I consider what these alternative factors might be. A significant but underappreciated element, I contend, can be found in the prominence of psychical research at the time. Early animal psychology was commonly affiliated with the occult, with practitioners from one area of research often dabbling in the other and several prominent animal researchers taking firm, public stances in favor of psychical investigation. The two shared

---

[25] See Coon (1992), De Sio and Marazia (2014), Lamont (2012), Sommer (2012; 2013), Wolffram (2006; 2009).

methodological affinities as well, adopting similar interpretive standards and a similar emphasis on observation and field experiment. That comparative psychology would be caught in the backlash was inevitable.

2.2 Psychology Old and New

Whether psychology admits of "revolutionary" analyses is controversial (Briskman, 1972). If any episode fits the narrative, though, it would be the transition witnessed by turn of the century comparative psychology. In a single generation, the field saw radical changes in how animals were studied and thought about. Research was pushed from the field or home into the lab, where animals were studied in aggregate using mazes, puzzle-boxes, and other apparatuses that helped distance subjects from experimenters. Darwin, Romanes, and even Lloyd Morgan published accounts stemming from their dogs' day-to-day activities. For Thorndike, Watson, and many of their generation, such anecdotes were useful only "for the purpose of illustrating how not to study and describe animal behavior" (Wheeler, 1904, p. 251). The way evidence was weighted changed too. Whereas earlier figures advocated choosing likeliest overall explanation or were, for ethical reasons, more comfortable erring on the side of over-attributing intellectual faculties (Mills, 1897), it became increasingly common for psychologists to admit "higher" processes only as a last resort (adopting a more stringently applied version of Lloyd Morgan's famed Canon).[26] According to Holmes, for instance:

---

[26] As Costall (1993) notes, Morgan's Canon was meant to work *with* the naturalism of Romanes, the standards of later years against it.

> The burden of proof is…placed upon those who contend for the superior endowments of the animal mind. What we want are not stories of performances which might have involved unusual intelligence, but records of achievements which *cannot be accomplished except by means of unusual intelligence*. In the latter case only are we justified in ascribing to the animal the mental attribute in question. (1916, p. 34, emphasis mine)

That this standard would "often fail to give to the animal full credit for the faculties it may really possess" was freely granted (p. 34), but this was simply the price of knowledge. Finally, and as so often happens in the wake of revolutionary upheavals, the monikers of the field underwent a process of revision. Whereas earlier texts bore names like *Mental Evolution in Animals* (Romanes, 1883) or *The Psychic Life of Micro-Organisms* (Binet, 1889), later authors overwhelmingly opted for a new, behavioral idiom. Leading publications bore names like the *Journal of Animal Behavior* and *Behavior Monographs.* In 1911, the heading of "comparative psychology" itself was replaced by "animal behavior" in the Psychological Index (Roback, 1920). A discipline once defined as the study of "all that pertains to the mind or soul of the animal kingdom" (Mills, 1898, p.17; Cf. Small, 1899, p.80) was now dominated by members who'd recoil at the suggestion.

Unsurprisingly, the discipline's view of its history largely followed suit. The field came to divide its existence into two eras or schools of thought, one "old" and one "new" (Holmes, 1916); one "anecdotal" and one "experimental" (Warden, 1928); one "false" and one "true" (Wheeler, 1904). Textbook introductions began labelling the work done prior to the late 1890s "subjective" and "anthropomorphic," backing their claims up with the most outlandish reports they could find in Brehm or Romanes (see, e.g., Washburn's [1917] discussion of the infamous "ant funeral"). Some went so far as to label it "pseudo-psychology" (Dunlap, 1922, p. 32; Wheeler, 1904). By contrast, later authors like Watson and Thorndike were praised as putting the discipline "on the road to objectivity" (Marquis and Hilgard, 1943, p. 6). "The anthropomorphic indulgences of the immediate post-Darwinian period," it was said, "had found their antidote in experimental investigation and theoretical skepticism" (Schneirla, 1958, p. 28). This interpretation would weaken with time, but large parts remain in place. The most recent edition of Hergenhahn's history presents animal research "before Thorndike" as prone to anthropomorphism, quoting a report relayed by Romanes of a parrot laughing "diabolically" after turning a plate of milk onto a cat (Henley, 2017, p. 348). Likewise, a recent introductory text by Wynne and Udell reports that it was only with Thorndike's 1898 study of associative learning that "animal psychology became a science" (2013, p. 7). The specific reasons for the field's change are rarely provided. Given the apparent progression and the stark contrast between the anecdotes quoted from Romanes (1883) and the neat puzzle-box experimentalism of Thorndike (1911), one is left to infer that its practitioners simply realized the founders' errors. The old psychology, bound as it was to anthropomorphism and anecdote, was a degenerating research program.

Beyond the standard worries about Whig history, there are several reasons to doubt the narrative. To start with, the early literature's reliance on and credence in anecdotal evidence has been much exaggerated. Lindsay, Mills, and Romanes all went out of their way to note that reports could be misleading and that every effort should be made to confirm them by independent observation and experimentation (more on this shortly). Indeed, the first of them flatly states in his *Mind in the Lower Animals* that "the value of anecdotes is apt to be greatly overestimated" (1880, p. 28) and that first-hand studies and experiment are "by far the best way to acquaint oneself with the phenomena" (p. 29). In line with this stated preference for non-anecdotal evidence, one finds a good deal of it in the field's first decades. Early developmental or "genetic" studies like those conducted by Mills (1898, part III) and Small (1899) with mammals or Pocock (1893) and the

Peckhams (1897) with arachnids involved months and sometimes years of lab work. At the same time, the literature contained a sizable number of experimentalists. Already in 1880, Lindsay could list 28 major experimental studies in the area (1880, pp. 32-4), many initiated as a means of testing anecdotal claims (e.g., ants' disposal of dead bodies; Lubbock, 1877). Researchers at the time investigated navigation (Darwin, 1873), tool use (Romanes, 1884, pp. 484-98), number recognition (Romanes, 1885; 1888, p. 58), and the effects of rearing animals in dark or monochromatic environments (Gates, 1895), to name only a few. Anecdotal evidence was certainly an important feature of early work, but other methods—including experimentation—were already an integral part of the discipline when Thorndike published his famous studies.

There is also a fair amount of exaggeration when it comes to early views on "anthropomorphism." For one, there has long been a tendency to quote authors selectively or out of context. This is particularly pronounced in the case of Romanes, who, as Greenwood (2016) has noted, chose to reproduce the reports he cited verbatim (1883, p. xi). Hence, most of the damning quotes used by later authors (e.g., "diabolically" laughing parrots) are not from Romanes but the sources from which he hopes to extract information. Indeed, the fact that these cases were relayed unedited may well reflect adherence to what Daston and Galison (2007, ch. 3) term mechanical objectivity: a vision of scientific practice common in Romanes' time that emphasized non-intervention on the part of the scientist. Just as microscopists and atlas makers of the time printed images with known artifacts and oddities, Romanes relayed anecdotal reports as he received them. That the reports themselves involved editorializing and were, for that reason, less than ideal was granted by all involved. To edit them, however, would have meant imposing himself on the reports and potentially withholding evidence that others may find relevant for assessing them. To attribute anthropomorphism to Romanes for reproducing the reports is, to some extent, to misunderstand the context. It's worth noting also that the period housed its own concerns about anthropomorphism. George Henry Lewes (philosopher and longtime partner of George Eliot) provides a case in point. In his writings on animal minds, Lewes cautioned heartily against "the fallacies of anthropomorphism" (1879, p. 124), showing no less distaste for metaphysical speculation than his 20th century counterparts. Crediting animals like bees with emotions, he argued, committed the considerable error of classifying by symptom rather than cause: a mechanical duck might accurately mimic its live counterpart, but move along without the slightest bit of feeling or intent (p. 126). Remarks cautioning against "anthropomorphic" preconceptions in dealings with animals can, in fact, be found in the work of a variety of 19th century authors (e.g., Herrick, 1895; vom Rath, 1891, p. 311). As with experiment, concern for anthropomorphism wasn't terribly new.

The central difference seems to be less about the importance placed on experiment or concern over anthropomorphism *per se* than the tone of the research and the extent to which psychologists recognized competing interests. While sensitive to anthropomorphism, the Darwinian naturalists placed equal or greater weight on evolutionary continuity and neuroanatomical homology (Lewes, 1879, pp. 140-2) or, in Mills (1897) case, the ethical implications of the evidential standards used. They conducted a host of experiments but did so in concert with other methods, in contexts other than the lab, and with a self-identified "sympathy" for their subjects (Lindsay, 1880). Like their Cognitive-era successors, psychologists gave their subjects names, examining their behavior in the field or at home. This ceased to be the case in the early-20th century, so much so, in fact, that it was hard for later psychologists to see prior work as scientific inquiry at all. Pleas for sympathy held little sway for those convinced that "all that has been claimed for the behavior of animals [must be] tried as by fire" (Wheeler, 1904, p. 251). Notes from the field or kennel were of little interest to a discipline that increasingly believed that an hypothesis was "of no value until it is made amenable to experimentation" (Dunlap, 1922, p. 32), and experiments testing whether Sally the chimp or Van the dog employed abstract concepts gained little recognition in a time where naming subjects might get one labelled an "insipid sentimentalist" (Carr, 1927). Both eras had an interest in controlled experiment, skepticism toward anecdotal evidence, and a concern for avoiding anthropomorphism, but something made those working in the early-20th century worried about personal engagement (naming, home-raising, professing sympathy, etc.) and the lack of certainty in a way that their predecessors were not. Framing the transition primarily as a move from anecdote and anthropomorphism to experiment overlooks this and risks mischaracterizing the intricate and, in many ways, more radical changes in sensibility accompanying the change.

This brings us to the final issue. Even if we bracket the standard narrative's assumptions about early research methods—even if the state of experiment and anecdote were just as they say—it wouldn't explain the outright bitterness of the transition. Disagreement over how best to conduct research had long been a feature of comparative work, with authors like Morgan (1894) and James (1890a, pp. 348-60) taking issue with evidence offered by figures like Darwin and Romanes. Near the time of the transition, however, these become far more hostile. Where early critics offered alternative interpretations or calls for further investigation of Romanes' claims, later work compared such studies to an "anatomy written from observations on dime-museum freaks" (Thorndike, 1911, p. 152); while skeptics like Morgan clearly saw themselves as part of the same endeavor as those they criticized, subsequent writings classified their opponents as "pseudopsychologist[s]" (Wheeler, 1904, p. 252) and traditional reports of animal intelligence as "fables" (Jastrow, 1906). Academic disputes can be heated, but the sheer number of insults and cordoning of the field's history—to say nothing of its reframing as the study of "behavior"—seem to go beyond this. Many fields experience methodological change, few maintain such a hostile vision of their past. At the very least, something is needed to account for this distancing, something that would make the methods appear threatening to the younger psychologists and not simply mistaken.

2.3 Historical Context

So, what does explain the extreme shift? As in any would-be revolution, there are a variety of factors. To date, at least two have gained widespread recognition by historians. The first is lack of resistance. For various personal and professional reasons, many figures sympathetic to the older approach ended up leaving the field (Richards, 1987, p. 505; Wilson, 2002). L. T. Hobhouse, whose *Mind in Evolution* sought to balance the new and old approaches, refocused his academic efforts on political theory and sociology after its 1901 publication; J. M. Baldwin, a champion of genetic psychology and admirer of Romanes (Baldwin, 1913, 86), had been exiled following a prostitution scandal (Richards, 1987, p. 496-503); and Lloyd Morgan, who was, like Hobhouse, heavily influenced by the work of older friends and colleagues, dedicated the better part of his last decades to the more philosophical questions.[27] Romanes, for his part, died in 1894 at the age of 46. Those hoping to maintain contact with the older Darwinian line were not wholly lacking in

---

[27] Again, Morgan's relation to the younger experimentalists is complex. Though famous for his stance on interpretation, it's clear he respected observational work and was deeply skeptical of Thorndike's version of experimentalism (Fitzpatrick and Goodrich, 2017).

support—Wesley Mills continued his work at McGill, while William McDougall remained an influential supporter throughout his career—but the loss was considerable. It may not explain the personal animosity of Thorndike or Watson, but this exodus does help explain their success. Had the figures listed above remained active, psychologists carrying on the older approach may have been better positioned. As it stood, however, the field's leadership was left in far less sympathetic hands (precisely whose, we'll see shortly).

Second, there were serious institutional pressures pushing for a more regimented, lab-based discipline, especially in the United States (Wilson, 2002; O'Donnell, 1985; Joncich, 1968, ch.8). In the time leading up to the period under discussion, the American university system had been in a state of transition. From a model focused on classical studies, colleges and universities increasingly turned toward matters deemed practical and scientific. Adapting elements from the German research university, institutions like Johns Hopkins and the University of Chicago established dedicated facilities for scientific research, with older institutions following suit. At the same time, land grant colleges like the University of California and Wisconsin were established for practical instruction in agriculture, teaching, and engineering. Unsurprisingly, given the discipline's dependence on these institutions, the psychology of the time reflects this trend. As O'Donnell notes (1985, ch.3), many of the first departments were founded with the aim of aiding teachers, and even at the budding research universities, considerable pressure was exerted on psychologists to yield practical results. Reading through materials of the era, one is repeatedly reminded that the nascent discipline is a science and that its fruits are, in the words of one psychologist, "not inferior to those offered by other experimental sciences."[28] Animal researchers,

---

[28] Wolfe to the Regents of the University of Nebraska, cited in Lundy (2006), p. 74.

it’s worth adding, were especially in need of such defenses, given the expenses of caring for their subjects and the release from teaching duties that such research often required (O’Donnell, 1985, p. 185). Seen in this light, it’s unsurprising that comparative psychologists would focus on topics like learning, which could be sold as contributing to pedagogy, and steer clear of anecdotes or at-home experiments, which might look like a hobby to administrators.

Even taking these forces into account, though, the transition seems a bit extreme. The loss of supporters makes sense of why certain lines of research declined and the ever-present need to demonstrate practical relevance helps explain the sudden appearance of learning curves in the journals. Yet, once again, the rhetoric surrounding the change, the overhaul of vocabulary, and the reframing of the discipline’s history suggest something more. Many disciplines underwent periods of institutionalization around the time, as the gentlemen naturalists of prior decades were gradually supplanted by professional scientists, but few so completely repudiated their history as animal psychology did. Few underwent so systematic an overhaul in terminology or so thorough a purging of case studies and observational methods. It’s one thing for up-and-coming scientists to criticize their forbearers or take up topics that fit with institutional priorities—Small’s (1901) experiments on learning and development model this as much as Thorndike’s—it’s quite another to label opponents pseudopsychologists or compare their subjects to dime-museum freaks.

One factor that may help to explain this severity, I argue, is the prominence of psychical research: the study of occult or “supernormal” phenomena like clairvoyance and trance states. Though rarely discussed in psychology departments today, such work was an important feature of the discipline in the late-19th and early-20th centuries. In its heyday, research dedicated to thought transference, trance states, premonitions, and the like were discussed in journals like *Mind* (Myers, 1893) and *Science* (Myers, 1888), as well as the discipline’s principal conferences. The

International Congress of Experimental Psychology, for instance, regularly incorporated such topics during its early years. Indeed, the 1892 meeting was overseen by none other than Henry Sidgwick, founder of the Society for Psychical Research (Myers and Sully, 1892). This prominence was not universally appreciated, however, as many psychologists believed the claims studied to be fraudulent and the very investigation of some issues “undignified” (Münsterberg, 1899, p.78). In fact, as work by Sommer (2012; 2013), Taves (2014), and others has shown, it was one of the most polarizing topics in the discipline.

To proponents like William James, the investigation of such “white crow” phenomena was the ultimate expression of empiricism. It signified a willingness to test and consider the evidence no matter what. For many of his colleagues, however, allowing such investigations to don the name of psychology risked lending credibility to known frauds and tarnishing the discipline’s image (Sommer, 2013). Prominent psychologists like James Cattell of Columbia, Joseph Jastrow of Wisconsin, and Hugo Münsterberg of Harvard spoke both publicly and privately about the dangers posed by the occult. Though diverse in their interests and views, each felt that the “pseudo-psychology” (Jastrow, 1930, p. 160) offered by psychical research posed an imminent threat and did what was in their power to prevent the psychologist’s public image from becoming that of a “spook-hunter or marvel-monger” (p. 197). Each wrote public denunciations in popular and professional outlets—comparing the relation between psychology and psychical research to that between astronomy and astrology (Cattell, 1888), attributing its phenomena to deception (Münsterberg, 1899; Sommer, 2012), and likening psychical researchers to medieval witch-burners (Jastrow, 1889, p. 77). Indeed, doing so was a public duty, as the research was “doing much to injure psychology” and James’ prominence was such that “he involve[d] other students of psychology in his opinions unless they protest[ed]” (Cattell, 1898, p. 641). Each did what they could behind the scenes, as well. Amid Cattell's correspondences, for instance, are Jastrow's thoughts on how to shape the dialogue surrounding psychical research and the "serious ill" it posed to psychology at upcoming conferences;[29] remarks about the "most decided pummeling" a recent book on hypnotism would require;[30] and complaints from Münsterberg about Harvard—an institution "in which [he had] fought for twenty years the cheap mysticism of the New England public"—inflicting "a deep humiliation…to the whole of psychology" by accepting money to research human immortality.[31] An association with the movement would be risky to say the least.

Now the idea that this dispute should touch upon *comparative* psychology is in some ways counterintuitive. To the present-day observer, questions about the thought processes of non-human animals seem quite orthogonal to those concerning, say, telepathy. Looking to the literature of the time, however, one finds the connection drawn explicitly and frequently. Dunlap's criticisms of anecdotal evidence are backed up with references to telepathy and premonitions (1922, p. 32, fn. 9), as are a dozen of Thorndike's attacks on "the anecdote school," whom he believed were "clearly in pretty much the same position as spiritualists" (1898, p. 107; Cf. 1911, pp. 15 and 241). In his classic introductory textbook on the study of animals, Watson argued that, when it came to the prodigious creatures offered by proponents of animal reasoning:

> The situation is exceedingly like that which appears in the investigation of so-called occult phenomena. Some new medium arises. Some prominent man visits the medium and becomes mystified. A scientific man, usually a physicist, is selected to investigate her. The physicist reports an elaborate series of tests which shows that the medium is not using

---

[29] Jastrow to Cattell, November 7, 1900

[30] Jastrow to Cattell, September 16

[31] Münsterberg to Cattell, January 29, 1913

> concealed wires, magnets, mirrors, or other physical equipment. The physicist himself may become "convinced." The mystery grows. Finally the psychologist makes the test and finds some simple trick which will account for the phenomenon.…In the same way, when a wonderful animal appears, zoologists, botanists, and physiologists are hastened to the scene. A commission is appointed and the mystery deepens. Usually when some man who is familiar with the methods of training animals and with animals' methods of responding is found, the explanation, while not necessarily simple, smacks not at all of the mysterious. (1914, p. 298)

The question is less *whether* these two domains were put together by psychologists but *why* they were and *how* this might have impacted the formation and reception of their project. I'll return to how in the next section, but for now let's focus on why.

2.4 Animals and Psychical Research

The most obvious reason for connecting the two is the substantial overlap one sees between the two domains. Culturally, both stemmed largely from the same liberal, middle- to upper-class circles and shared many of the same values (see Moore, 1977, ch.2). Both were associated with the animal welfare movement and natural conservation efforts, for example. Examining spiritualist periodicals of the time, moreover, one finds that the community maintained an abiding interest in animal minds. For instance, the leading newspaper *the Spiritualist* reprinted the entirety of Romanes' BAAS lecture on "Animal Intelligence and Comparative Psychology" as well as early experiments by Lubbock on the emotions of ants (1878, pp. 91-2). Ten and twenty years later, one finds similar coverage in outlets like *Light* and the *Annals of Psychical Science* dealing with animals' awareness of death (Cummings, 1888) and sensitivity to music (Fotherby, 1908). What's more, the interest was mutual. Naturalists and comparative psychologists were among the earliest supporters of research into the occult. In the French-speaking world, for instance, prominent figures like Richet, Flournoy, and Claparède published in both areas. In England, meanwhile, writers like Romanes and the ornithologist Edmund Selous supported the movement through interactions with the Society for Psychical Research. Indeed, the former had been among the SPR's earliest members (Bennett, 1903, pp. 5-6) and a firsthand reporter of paranormal phenomena, having witnessed an apparition at the time of his sister's death (Romanes, 1890). Though frequently skeptical of spiritualism's popular expressions, he showed no hesitation in calling out those closed-minded scientists who refused to even attend a séance (1895, p. 109).[32]

A still more striking illustration of the domains' overlap comes from the great deal of psychical or otherwise supernormal research directly concerning animals. Texts discussing psychic powers of animals can be traced back some time (see Jones, 1861, p. 176), but their heyday is likely the period leading up to and immediately following the turn of the century. Both psychical and comparative research were growing, and many believed that non-human animals were or might be more sensitive to the supernormal, possessing faculties that remained mostly subconscious in humans (Bozzano, 1905, p. 119). Animals were said to have witnessed and appeared as apparitions, to respond to locations with occult ties, and to experience premonitions of natural disasters (Bozzano, 1905; Hudson 1893, ch. 6, 9). A small literature had also emerged on such abnormal or supernormal conditions as hypnosis and dual- or complex-personality, urged

---

[32] Though less strongly associated with comparative *psychology*, it bears noting that the naturalist and co-discoverer of natural selection A. R. Wallace was among the most prominent supporters of popular spiritualism in the era just prior to that under discussion (see Slotten, 2004, ch. 10 and 14).

onward by prominent naturalists like Mills (1897) and Selous (1905, pp. 289-95). Perhaps the most important area of study, however, concerned thought-transference. Mental connections were believed to account for coordinated behaviors like those observed in flocks of birds—a thesis defended by Selous (1901, ch. 8; 1905, p. 77) and the American naturalist John Burroughs (1905, pp. 96-8)—while human-animal connections were reported across the U.S. and Europe. In Germany, the Elberfeld estate of Karl Krall housed Equine psychics like Zarif and Clever Hans, as well as Basso the numerate and telepathic chimpanzee (Candland, 1995, ch.5-6; De Sio and Marazia, 2014); in England lived Bob, a "thought-reading dog" who could identify locations on a map and perform simple arithmetic when his trainer simply thought the question ("A Thought-Reading Dog," 1903; figure 1); and in America lived Dodgerfield, a terrier whose intellectual and telepathic prowess earned him an 1897 meeting with the Boston Scientific Society ("The World of Psychics," 1897). One could even find how-to manuals for psychical pet training (Fitt, 1890). Nor were such ideas known only to those in the spiritualist and psychical research communities. The Elberfeld horses were discussed in *American Journal of Psychology* (Sanford, 1914), *Psychological Bulletin* (Porter, 1914), and *Archives de Psychologie* (Claparède, 1912), while various American counterparts found mention in outlets like the *American Naturalist* (Henshaw, 1897, p. 561) and *North American Review* (Quakenbos, 1912). In short, animal research—especially those forms dealing with "higher" faculties—would've had a strong association with the occult.

Figure 2. Bob the telepathic dog. Image from the Biodiversity Heritage Library. Contributed by Harvard University, Museum of Comparative Zoology, Ernst Mayr Library.

Finally, even if there hadn’t been such personal connections between animal and psychical research, the programmes were remarkably similar in methods. For one, the era’s psychical researchers drew heavily on testimony, making use of historical records, letters to the SPR, and surveys, among other sources. The monumental *Phantasms of the Living,* to cite just one example, brought together literally hundreds of reports, with entries coming from newspapers, letters, and academic journals around the world. As with Romanes’ texts on animal intelligence, these were weighted to favor first-hand testimony from known correspondents and, in the interests of transparency, largely reproduced verbatim. The operative idea in both cases was that the collation of such reports constituted a first Baconian step toward generalization and an aid for identifying new areas of research. Similarities could be found in experimental practices, as well. Though psychical research would eventually pursue laboratory methods (including in animal psychology, see De Sio & Marazia, 2014), work in the period under discussion was typically based in the field, viz. the parlor or séance room. Few dedicated labs existed at the time, and in any event, at-home settings were the most comfortable for the mediums. SPR experiments would often involve a prominent researcher or set of researchers (James, Myers, etc.) attending a séance, gathering background information, weighing alternative hypotheses, and so forth. Like Romanes' field experiments, investigations would frequently be conducted using controlled subject-experimenter interactions and convenient objects like playing cards (Gurney et al., 1886, pp. 31-5). Efforts were made by both groups to remain impartial, though the typical researcher was open to the idea that the phenomena studied may be legitimate (Oppenheim, 1985, ch. 4). There could be a trick, but the horse might really be counting or the medium sensing the experimenter's thoughts. Finally, many psychical researchers were, like the naturalists who reported on pets and zoo animals, on friendly terms with their subjects. James, for instance, once declared that he "should be willing now to stake as much money on [the medium] Mrs. Piper's honesty as on that of anyone I know" (1890b, p. 654). Lastly, and as I'll discuss later, both felt their opponents were far too hostile, deciding the issue *a priori* or insisting upon unreasonably conservative interpretative standards.

In considering how this association between comparative and psychical research might have shaped the former, there are at least two potential impacts that stand out. To start with, the animosities that had been stewing in the latter domain could, given the association, be expected to transfer. On first reading, the words of Thorndike, Wheeler, and their colleagues sound exceptionally hostile. Relative to what had been said of psychical research, however, they are rather unremarkable. Indeed, the words and phrases are often identical. Whether this stems from genuine feeling or a rhetorical savvy on the part of the authors is difficult to say, but in either case, the fact that they engaged in accusations of storytelling and pseudopsychology is less surprising. Second, the methodological parallels between the areas of study would have lent the younger authors' arguments a measure of force that, in the absence of hostility toward psychical research, may have been lacking. If anecdotal evidence and personal connections to one's subject were rejected in the one case, it would be hard to avoid doing so in the other. This fact would be sharply underscored by the prominence of thinking animals in the spiritualist press, the association between comparative psychology and psychical research in the persons of Romanes, Selous, and others, and the ever-growing literature on psychic animals. It simply remained for someone to seize the polemical opportunity, taking us from why to how.

2.5 Two Lines of Attack

In considering how the anti-spiritualist climate may have influenced comparative psychology, there are two interrelated factors one needs to consider. The first is broadly epistemological. The success of the behavioral turn can be traced in part to proponents' effective use of methodological connections noted in the previous section. Drawing on a pre-existing pattern of argument from debates over psychical research, critics of the older animal literature were able to leverage fears about the discipline's future and the supposed dangers of using anecdotes, field observation, and less conservative interpretive standards in establishing psychological claims. These concerns may have carried some force independently of the associations and similarities just mentioned, but the fact that they came at a time when psychology was thought to be under threat and were directed at many of the same individuals doing the alleged threatening doubtlessly helped. The second is more political in flavor, having to do with the ability of experimentalists to secure institutional support from established anti-spiritualist psychologists. As I mentioned earlier, psychology housed many figures hostile to psychical research and eager to reform the field. By framing their case in terms drawn from earlier debates, it seems, the revolutionaries were able to obtain extensive support and publicity for their work. Favors were bestowed liberally on those adopting the new approach, and this appears to have given them an advantage in the ongoing disputes. Each shall be discussed in turn.

2.5.1 The Methodological Critique

The outline of the main arguments can, I suspect, be inferred largely from the foregoing sections. The risks of "the anecdotal method" and virtues of lab-work have been standard fare in textbooks for over a century and can likely be listed by even the strongest proponent of naturalistic observation. It's nevertheless worth considering some of the specifics, as the original sources reveal connections to the psychical research literature since lost. The core methodological critique can be found in Thorndike's 1898 dissertation—commonly cited as the beginning of "scientific" or "experimental" animal psychology—and consists in a few key claims. The first was that anecdotes and personal testimony were inherently unscientific. For one, the events recorded were subject to errors of memory, generally prejudiced in favor of their subjects (see below), and almost always selected for being noteworthy rather than representative. As Thorndike put it, "the anecdotes give really the *abnormal* or *supernormal* psychology of animals" (p. 25), one emphasizing "the unusual and marvelous" behaviors to the neglect of the everyday and "stupid" ones. Second, even if the reports were accurate, there were issues of generalizability. Reports generally dealt with single cases and often relied on such fortuitous combinations of events as to render replicability, the very lifeblood of science, an impossibility. An accurate report might tell you that something happened, but not whether it holds generally; in short, anecdotes didn't make laws (p. 25). Turning from the second to the first-hand, the situation little improved. The field observations and simple experiments characteristic of prior work, he noted, took place under far from ideal conditions. Things like the animal's previous experience were frequently left open (p. 25), and the possibility of inadvertent experimenter or trainer cuing were uncontrolled—in many cases, the animal need only look to the experimenter's face to see the answer (p. 125). Indeed, Thorndike claimed, "far greater care has been given by the Psychical Research Society to establishing the criteria of authenticity [and] to insuring good observation" than earlier psychologists (p. 152). Finally, observers were far too involved with their subjects. Reports were made on pets whom the observer clearly cared for, and as a rule, human observers seemed disposed to see the marvelous in nature (even those who doubted their reason could not help but find something else to celebrate; p. 23). These facts inevitably shifted work toward certain projects over others (p. 23) and prejudiced their interpretation; just as "men of the utmost scientific skill have failed to prove good observers in the field of spiritualistic phenomena," Thorndike warned, "so biologists and psychologists before the pet terrier or hunted fox often become, like Samson shorn" (p. 25).

"To remedy these defects," Thorndike claimed, "experiment must be substituted for observation and the collection of anecdotes" (p. 26). Better observers and more judicious sorting of testimonies wasn't enough. At the same time, psychologists would have to stop treating the animals so reverently. Astronomers and bacteriologists, he argued, didn't praise or marvel at their subjects like the doting naturalists of the anecdote school (p. 23), and if comparative psychology was to be a science, it would need to follow their lead. The innate appeal of the "intelligent and unusual" couldn't be allowed to overtake the "stupid and normal" as it had so often done in comparative and psychical research both (p. 25). His own experiments were an illustration of this novel approach. His animals were studied in aggregate, ameliorating concerns about abnormality; their conditions were uniform and repeated over many trials, guaranteeing some reliability; and critically, the experimenter was removed from the equation as much as possible. Experiments examined animals' ability to escape from puzzle boxes (figure 2) rather than respond to requests, involving the experimenter only as an observer. Thus, Thorndike assured his readers, the studies were "uninfluenced by *any* personal factor in any way" (p. 28, emphasis his). This went doubly for the issue of attachment. Thorndike didn't name his animals in print or live alongside them, as prior naturalists had done, but instead housed them in labs (i.e., in attics or cellars) and assigned them labels like "Cat 10." Meanwhile, his emphasis on animal "stupidity" over "intelligence" (p. 22), use of cages, and practice of keeping his subjects in a state of "utter hunger" (p. 27) removed any concerns readers might have had about sentimentality. In developing his own laws of behavior, Thorndike presented himself as no more attached to his subjects than a bench chemist was to an organic compound.

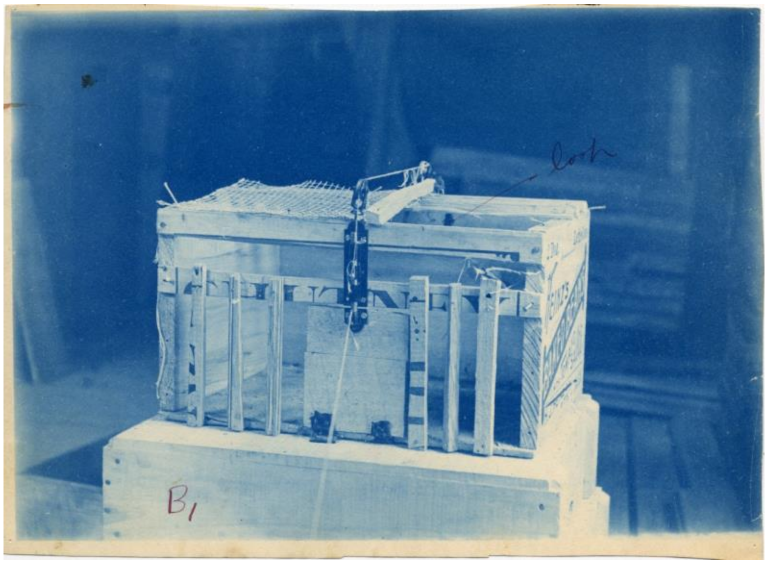

Figure 3. Thorndike Puzzle from Robert Mearns Yerkes papers, 1822-1985. Manuscripts and Archives, Yale University Library.

The challenges weren’t without force, but neither were they novel. That conditions in the field weren’t ideal, that critical distance was needed, and that experimental interactions needed to be controlled were, as we saw in section 2, familiar to earlier authors. The difference was that they believed such problems could be overcome and that the measures adopted by Thorndike and his fellows came at too high a price. Anthropomorphism could be checked by physiological considerations; inadvertent cues could be controlled by using standardized stimuli (Romanes, 1885); and the risks associated with anecdotal evidence could be ameliorated by encouraging critical scrutiny and prioritizing reports made by trained naturalists or communications with photographic evidence (e.g., Witmer, 1910). Meanwhile, the practice of dismissing anecdotal or purely observational evidence in favor of lab-work risked ignoring key evidence (Long, 1904) and putting too much confidence in results derived from artificial conditions and interpreted by experimenters palpably hostile to the idea of animal reason (Mills, 1904). The “utter hunger” of Thorndike’s animals and his use of tight cages (20x15x12 inches for cats) seemed likely to shift the animal from a state of reflection to a state of panic, and even if the animals hadn’t been so enclosed, it was argued, mechanical puzzles were hardly a domain where their intelligence would be expected to manifest (pulleys, trap-doors, and the like being foreign to their lives and evolutionary habitats; Hobhouse, 1901; Mills, 1904, pp. 747-50). Finally, when it came to interpretation, the conservative standards implicit in his project and pushed explicitly by Wheeler, Holmes, and others risked bypassing better-evidenced explanations in favor of “lower” ones. The last-resort standard may lead one to fewer mistaken attributions of “higher” psychological activities, but this seemed like a misplaced value. “Why should we bind ourselves by a hard and fast rule like this one” observers could ask, “[i]s it not the truth at which we wish to get?” (Mills, 1898, p. 11).

As with Thorndike's own criticisms, these concerns are not *prima facie* implausible (interestingly, many would re-emerge when fieldwork and live-in experiments became popular decades later). The presence of psychical research rather changed the dynamic, though. As we saw in the previous section, comparative and psychical research overlapped a great deal in their methods, opening the two to the same objections. In fact, if one looks to the years leading up to Thorndike's dissertation, one finds the exact same arguments and objections populating the literature on psychical research. The testimonies collected by the SPR and similar organs were dismissed as "unmethodically gathered miracles" (Münsterberg, 1889, quoted in Sommer, 2013) rife with "mal-observation and faulty description" (Jastrow, 1889), while the researchers themselves were presented as unduly credulous, personally attached to their subjects (as James was to Piper), and emotionally invested in their study's outcome (Cattell, 1898). Critics argued that psychical researchers needed to exhaust all "normal" explanations before turning to notions like thought-transference (Jastrow, 1889) and declared that to believe on insufficient grounds was to "catch a stain which can never be wiped away" (Clifford, 1884, p. 27). What's more, they regularly characterized the mesmerized subjects and entranced mediums studied by the SPR as "pathological" (Münsterberg, 1899, p. 67; Wundt, 1896, pp. 336-337). At best, it was suggested, they were unwittingly misleading investigators; at worst, they were complicit in fraud. Finally, and most significantly, the observations and experiments on which psychical research relied were said to involve too many uncontrolled variables to offer evidence for the researchers' claims (Jastrow, 1889; Münsterberg, 1889; 1899). The events seen in a dim parlor could be caused by any number of factors, as could the success of a purported telepath. Strikingly, the very mechanisms by which mediums were said to operate were the same as those later attributed to clever animals, viz. subtle cues produced by their investigators. As early as 1877 doubters suggested that mediums and mentalists could be using changes in the hands and posture of their patrons as a guide, a practice referred to as “muscle-reading” (Beard, 1882). Over time, this would extend across a multiplicity of channels: unwitting sitters might blush, shift their breathing, look away, or cue the medium in any number of ways (Münsterberg, 1899, p. 69). It was precisely this trick that Thorndike believed to underwrite the performances of a numerate chimpanzee reported by Romanes and an unnamed psychic dog (likely Dodgerfield; 1899, p. 88); it was also the method that anti-spiritualist Carl Stumpf and his students used to account for Clever Hans (Pfungst, 1907), providing psychologists with a cautionary tale lasting generations (Candland, 1995, ch. 5-6).

As was the case for their peers in comparative psychology, the psychical researchers were not without replies. Here again, the form of the argument is startlingly similar. For one, they argued, the doubters were at least as emotionally invested in the situation as they were. While the psychical researchers generally took care to weigh alternative explanations and isolate sources of error, they claimed, the same could not be said of the hardliners (James, 1986). Critics regularly denigrated the subjects and practitioners of psychical research, refusing to even investigate the phenomena in question. Rather than consider the claims disinterestedly, they insisted that “normal” explanations like fraud or cuing be preferred to “supernormal” interpretations regardless of the specifics of the case. Yet appeals of this sort were available *in abstracto* for literally any result, and those citing them rarely took the time to flesh out how the feat would be accomplished in controlled studies of the kind undertaken by the SPR (James, 1898; Gurney et al., 1886, p. 18). Muscle-reading, for instance, could account for many reports, but some of the information guessed by the mentalists seemed far too exact to be conveyed through bodily cues. When it came to *a priori* commitments, Romanes claimed, these “scientific men” were “quite as dogmatic as the straightest sect of theologians” (1895, p. 109). Second, proponents argued, critics greatly exaggerated the risks associated with testimony and observation. Eyewitness records and participant observations were fallible, of course, but they were a source of information nonetheless. Many psychical phenomena (e.g. visions of a distant loved one's death) were ill-suited to lab study, and in any event, dismissing the "mass of testimony, or apparent testimony" of one's peers was hardly balanced scholarship.[33] At best, refusing to consider reports simply because they *might* contain errors placed the avoidance of error over the search for truth—aims that, as James (1896) famously noted, frequently come into conflict. At worst, it did little more than exclude evidence the critic found discomforting. Finally, it was sometimes argued, the conditions imposed by hostile or 'debunking' experiments were manifestly adverse to mediumship. Stress, anxiety, and even the presence of uncooperative or unsympathetic members in a séance were commonly thought to interfere with the medium's efforts (Wallis and Wallis, 1903, p. 109), so it was hardly surprising that some should fail to reproduce performances under the eyes of hostile observers. In time, the exact same line of response would be made on behalf of Clever Hans and his fellow *Denkende Tiere* (Candland, 1995, ch. 5-6).

Even if Thorndike hadn't repeatedly invoked the occult in his writing, the prominence of spiritualism and psychical research would have been enough for readers to grasp the underlying argument. When he claimed that comparative psychologists were relying on "abnormal" subjects or assembling an anatomy based on "dime-museum freaks," readers could recall similar remarks by Wundt on hypnotism and Münsterberg on trance mediums; when he pushed for the priority of "normal" explanations before "supernormal" ones, they could recognize the arguments of Jastrow; and when he wrote of scientific men being "like Sampson shorn" before the psychic or pet terrier,

---

[33] James to Stumpf, January 1, 1886

they could easily call to mind the James-Piper relationship or, if they'd been in Boston around the time Thorndike was, the public display of the psychic terrier Dodgerfield. If researchers allowed the standards and methods of comparative psychology to stand, one could infer, it would leave the door open to all manner of spirits and unseen forces. Intellectual consistency could not let the sifting of anecdotes, use of field observation, professed sympathy with subjects, or use of less-rigid interpretive standards remain in one case and fall in the other. While proponents of the earlier approach had replies to Thorndike's specific objections, the wider associations were difficult to shake, as the replies themselves were frequently prefigured in the psychical research or spiritualist literature. (It little helped that the field-based approach was explicitly endorsed in outlets like the *Theosophical Review* [Maguire, 1907]). To those worried about the discipline's image, including many of the most powerful figures in psychology, the choice was clear: however strict the measures, they were needed to preserve the field.

2.5.2 Cementing the Approach

This brings us the structural aspect of the story. The implicit and explicit comparisons drawn between early comparative and psychical research mattered not just because they persuaded average readers that change was necessary but because they gained their authors the backing of leading figures, particularly in the United States. The anti-spiritualists discussed above were exceptionally supportive of their younger colleagues, providing them with a range of institutional supports the absence of which would have severely hampered the growth of the movement (jobs, labs, etc.). Their reason for doing so cannot be determined exactly but tensions over the occult seem like a likely contributor. For one, some draw the comparison between traditional comparative work and psychical research explicitly (see below). Second, the methodological critiques being used by the revolutionaries were, in a sense, their own. The difference was simply one of targets—and even this might be disputed, given all the work on animals conducted by psychical researchers. Finally, there is the general tendency in times of extreme polarization for groups to circle the wagons, protecting those who share their views on the critical issue and opposing those who do not. Older figures like Jastrow and Cattell were vehemently opposed to spiritualism and psychical research, and we know from the foregoing sections how the traditional and lab-centered comparative psychology tended to split on the matter. Given their interest in reshaping the field, there's reason to think they'd see the young animal researchers as a good investment.

At least three forms of support are worth noting. The first and most obvious was simple promotion. Supportive discussions of the new programme came in publications both popular and technical, with the menace of psychical research constantly lurking in the background. Münsterberg, for instance, credited the movement with turning animal psychology into an "exact science" and liberating it from the "gossip" and "queer observations" of the past (1914, p. 223), while Jastrow denigrated research on exceptionally intelligent animals as little more than make-believe in his "Fact and Fable in Animal Psychology" (1906), styled after his anti-spiritualist "Fact and Fable in Psychology" (1900). Indeed, the latter author had thrown his support behind the new approach as early as 1900, taking time in his APA presidential address to highlight what he saw as the yoked problems of human and animal intelligence testing and the role of rigorous measurement in the fight against "haunted houses and thought-transference" (1901, p. 16). As he later stated:

> I felt it incumbent to proclaim that a psychologist was not primarily a spook-hunter nor was his chief diet tales of psychic long-distance communication and trance revelations….I defended and selected as of greatest significance and promise the two movements then

> struggling for recognition: the precise study of animal behavior as the primer of human behavior, and applied psychology in its salient ramifications. (Jastrow, 1935, p. 98)

His characterization of the movement as “struggling for recognition” is likely an overstatement—Thorndike’s dissertation had caused something of a stir in 1898—but it’s remarkable nonetheless that the effort would be showcased on such a prominent platform. Even if the young experimentalists had gained coverage earlier, Jastrow’s speech and later articles would constitute a powerful signal of support.

This brings us to a second and more substantial support. Older psychologists regularly backed their colleagues when it came to space and funding, ensuring that the promise signaled by speeches and articles were fulfilled. We know, for instance, that during his tenure as head of the Harvard lab Münsterberg set aside increasingly large sections for animal research. When Thorndike first began his animal studies in 1896, room was so scarce that the experiments were run from William James’ cellar (Joncich, 1968, p. 87); by 1913, animal labs occupied an entire floor of Emerson hall (Yerkes, 1914), their output constituting roughly half of the *Harvard Psychological Studies* (Münsterberg, 1915). Cattell was similarly supportive at Columbia, securing opportunities for Thorndike and Watson both. He was credited with having gotten Thorndike invited to the prestigious Woods Hole biological laboratories, for instance, and for gaining him a position at the Columbia Teacher’s College.[34] A few years later, he was similarly generous in offering Watson—who, unlike Thorndike, hadn’t been his student—repeated summer lectureships and research positions. “I have always heard that you were a man who did not allow friendship to enter into your business relations,” Watson joked upon the third invite, “but I fear

---

[34] Thorndike to Cattell, February 1899

that you have been misjudged. I am sure the two years I have taught in Columbia have not been so valuable to the university as to call for this third offer."[35] These lectureships ensured that Columbia students were receiving a hardline experimentalist education, and more importantly, gave Watson room to work-out his ideas. It was from these lectures that he developed his landmark "Psychology as the Behaviorist Views It" and *Behavior: An Introduction to Comparative Psychology*, texts that did as much as any to cement the trend. Cattell continued to channel money and students to Watson through 1916 and '17;[36] when the latter eventually published his *Psychology from the Standpoint of a Behaviorist*, it was appropriately dedicated to Cattell.

Finally, and perhaps most importantly, they seem to have been given preference in publication. Cattell appears to have been especially important in this regard. As the editor of the *Psychological Review* from 1894 to 1904, head of the *Popular Science Monthly* from 1900, owner of the *American Naturalist* from 1908, and the editor of *Science* from 1896 till his death, Cattell "presided over the destinies of psychology for generations" (Pillsbury, 1949, p. 1). No small measure of this power, it seems, was directed toward the advancement of the lab-focused psychologists. As an editor, Cattell solicited reviews and published materials from those he felt were doing the most scientific work. This naturally included work by methodologically conservative biologists like William Wheeler and H.S. Jennings and psychologists like Thorndike, whose career-launching study of association was released in the *Psychological Review Monographs*. To give a sense of the balance between the new and old style of animal research in Cattell's journals, we may note that between 1896 and 1910, *Science* published at least twelve

---

[35] Watson to Cattell, November 22, 1910

[36] Watson to Cattell, May 16, 1916; April 26, 1917; May 11, 1917

contributions by Wheeler but that in the same period—and in striking contrast with the similarly positioned *Nature*[37]— featured virtually no coverage of old-style work (Turner's [1909] letter on snake behavior being a potential exception). Purely observational work was rare; anecdotes were unheard of. In the case of the *American Naturalist* and *Popular Science Monthly*, moreover, we find outlets previously open to alternative approaches closing their doors post-acquisition. Before Cattell's 1908 purchase, the former journal had published a mixture of perspectives, including reviews sympathetic to Mills' observational work (MacDougall, 1899a, 1899b) and animal hypnotism (G.H.P., 1899). Afterward, it switched to work "on a high plane,"[38] focusing on "strictly scientific" investigations (Jennings, 1908a; 1908b; 1908c) and advancing the "radically experimental" view that a question for which feasible crucial experiments were lacking was "one with which science can not deal" (Jennings, 1913, p. 349). The same goes for the *Popular Science Monthly*. Prior to Cattell's 1900 acquisition, the journal had published dozens of articles on topics like laughing in dogs (deRochas, 1896), "animal speech" (Evans, 1893), and "education" across the animal kingdom (Letourneau, 1898); Mills alone had published at least seven articles with it. After 1900, however, articles focused increasingly on "behavior" over intellect (Thorndike, 1901a; 1901b; Wheeler, 1906; Yerkes, 1901; but see Herrick, 1910), dismissing reports of exceptional animal accomplishments as "fables" comparable to those offered by the spiritualists (Jastrow, 1906). As time went on, the traditionalists had fewer and fewer outlets to which they could turn. Indeed, the only major U.S. journal featuring such work was the *American Journal of Psychology*,

---

[37] *Nature* coverage included a letter on the "histrionic capacity of Grey Parrots" (Fearon, 1901) and numerous reports of animal intelligence (J.E.A.T., 1904; Galloway, 1906; M.N.W., 1912).

[38] Watson to Cattell, July 4th, 1908

a publication that Cattell had unsuccessfully tried to acquire in the early aughts. (The AJP also housed discussions of psychical research, including with animals; e.g., Sanford, 1914).

Thus, in addition to the rhetorically difficult position they were in *vis a vis* Thorndike and the current mood in psychology, those incorporating testimonials and naturalistic observation would have faced an uphill battle in the academy. While their ideological opponents had a strong and vocal backing, the traditionalists had, as was noted earlier, comparatively little support. There were a few sympathizers—E.C. Sanford and his students at Clark (e.g., Kline, 1899) being the most conspicuous—but the older psychology had largely been an occupation of independent naturalists. Darwin, Romanes, and Lubbock all worked outside the academy, and many of their early twentieth-century followers did too. Such figures could register concerns with the path animal psychology was taking, but few had the kind of clout or institutional power wielded by Cattell or Münsterberg. In a world increasingly dependent on universities and bodies like the AAAS, they were at a disadvantage.

2.6 Conclusion: Exorcizing Ghosts

If the arguments presented in the foregoing sections are correct, then the behavioral turn in comparative psychology (and perhaps psychology more generally) owes a certain debt to psychical research and the reactions it engendered. Popular and scientific investigations of the occult contributed to increasing polarization in academic psychology. Eager to distance themselves from the "ghost hunters," many adopted a rhetoric hostile to the use of testimony or anecdote and deeply concerned with the possibility of holding erroneous beliefs—or being duped. Animal psychology was one domain where this found powerful expression. The field's association with psychical research and strong methodological similarities with that programme left it vulnerable to the same challenges and hostilities, and as we have seen, these were used to powerful effect by younger psychologists and biologists advocating the strictly lab-based and zealously unsentimental approach that came to dominate the field. In making their case and establishing their alternative vision, they were aided by established anti-spiritualists—given positions and space in university facilities, celebrated in leading disciplinary meetings, and printed in some of the most widely circulated scientific journals. In the hands of Thorndike, Jennings, and others, the field of animal behavior was made to display all those virtues supposedly absent from occult investigations.

How one feels about these efforts will likely depend on one's views on the proper domain and methods of scientific inquiry. Those sympathetic to the early approach may see the episode as a missed opportunity. Had the discipline been less polarized at the time, 20th century psychology might have developed a far more robust tradition of field work or a more sympathetic vision of its subjects. Those who reshaped the field may seem unjustified in their efforts and manipulative in their repeated invocations of popular spiritualism. By contrast, those in favor of the standards that emerged during this time may see the efforts as a service to psychology. Thorndike may be lauded—not, perhaps, for introducing experiment to the field but for seizing the moment and acting to correct a situation spiraling out of control. Which version of the narrative holds the greater truth is beyond the scope of the current work. Whichever way we choose to see the transition, however, it benefits us to do so in light of a fuller understanding of its institutional and social context, cleansed of the aura of inevitability that so often surrounds narratives of revolutionary change.

3.0 Third Paradigm: Cognitivism and the Case of the Cognitive Map

3.1 Introduction

Today's comparative psychology is no longer that of Thorndike and Watson, of course. In the intervening years, we have seen a shift not unlike that which accompanied the behavioral turn. Beginning in the thirties or forties and gaining steam through the end of the last century, a new set of terms and topics began to emerge. Leading books, journals, and conferences came to bear the name "cognitive," housing discussions that would've made the prior generation wince. Where talk of mental states was once taboo, researchers now ask unabashedly whether rats can empathize, elephants have a sense of self, or chimpanzees are capable of attributing mental states to others. The move is generally celebrated among practicing psychologists, for whom the older norms felt like fetters to inquiry. For all its success, however, the ascendant cognitive approach has had an exceptionally troubled reign. The questions that it helped usher into the field—questions about the internal structures and processes "behind" behavior—have remained subjects of intense controversy. Time and again, results promising to settle questions about memory, reasoning, and the like have been met with conflicting evidence or alternative explanations (Shettleworth, 2010). Few would choose to return to the days when such questions were anathema, but as time has gone on, psychologists have grown weary of debates widely seen as a "muddle" (Osvath, 2016, p. 5), "more sterile than fruitful" (Beckers, De Houver, and Dwyer, 2016, p. 187), or simply "in trouble" (Heyes, 2014, p. 313).

The question is why? What is it that makes animal cognition so prone to these difficulties, and what, if anything, can be done about it? Numerous answers have been offered over the years, ranging from comparatively minor experimental adjustments to wholescale rethinking of the problems at issue. To date, however, no explanation has gained widespread acceptance or proven capable of resolving the problem. In the following, I mean to present an alternative diagnosis. While most recommend more careful *thinking*, I suspect the problem is material, stemming from the field's instrumentation rather than its methods of theory construction or analysis. As a science, animal cognition relies on behavior as its main source of information on upstream mental processes, with subjects' motor control pathways serving as the material channel or "instrument" by which this cognitive information is relayed. Considered as a cognitive measurement system, however, these pathways fall short in a number of ways. While they serve their evolutionary functions of selecting and executing actions quite well, they carry too little information and are surrounded by too many unknowns to serve the *epistemological* functions demanded of them. Behavior is an insufficient measure of cognition for the questions being asked, and better theories, tasks, and data analysis techniques are unlikely to circumvent this problem.

I begin with a review of the issue, distinguishing two central challenges in the literature on animal cognition and examining the principal factors thought to contribute to the controversy. While they may serve to explain some of the difficulties, I argue, there is much that these proposals leave out. From there, I develop the outlines of an alternative, instrument-focused analysis of the controversies. I distinguish three central dimensions of measurement applicable across scientific disciplines, considering how each applies in the case of animal minds. Finally, I provide a detailed case study of a paradigmatic controversy from the discipline. Examining the history of 20th and early-21st century debates over "cognitive maps," I show how instrumental shortcomings yielded otherwise unexplained patterns of conflict in animal cognition and other controversy-prone disciplines. Those seeking to understand and address scientific conflict, I conclude, cannot afford to overlook the tools by which scientists act and gather information on their targets of study.

3.2 The Cognitive Challenge

Animal cognition, as the term is used here, designates a branch of comparative psychology focused on the internal or "cognitive" mechanisms underwriting capacities like memory, learning, and decision-making in non-human animals. Historically, it emerges around the mid-20th century, as the primarily behavioral theorizing that characterized much previous work in ethology and behaviorist psychology transitioned to a more mentalistic terminology. Methodologically, however, it retains a number of features in common with its predecessors, including certain standards of evidence and experimental design. Most relevant for the discussion here, the field maintains a strong commitment to behavioral techniques, and when combined with its theoretical orientation, this gives it a distinctive research profile, viz. the use of behavioral measures or "criteria" to individuate specific mental processes or representations. It is by observing patterns of action in the field and testing them in the lab that the student of animal cognition seeks to determine whether or in what sense elephants have a sense of self or scrub-jays have episodic memory.

In its time, the discipline has had no small number of successes, particularly when it comes to establishing behaviorally defined *capacities*, or the basic ability of an animal to solve some problem. The discovery of chimps' use of tools and octopuses' ability to solve multi-stage puzzles, for instance, have dramatically altered both scientific and public perceptions of these animals. Whatever structures lay within the minds of these animals, we can infer, must at the very least be capable of performing these feats. When it comes to characterizing these mechanisms and processes in a more positive way, however, progress has been exceptionally hard coming. While it's widely recognized that chimpanzees keep and modify tools in advance of their use (Sanz et al., 2009), it remains controversial whether their doing so results from the envisioning of future problems, a sophisticated form of habit learning, some task-specific drive, or a combination of factors (Suddendorf and Corballis, 2010; Osvath and Martin-Ordas, 2014). Likewise, while it's clear that certain animals "pass" the famed mirror self-recognition test, it's hotly disputed whether this success reflects an abstract sense of self, bodily awareness, or something else (Kohda et al., 2019). Indeed, such longstanding disputes can be found in nearly every domain of cognitive study, including causal reasoning (Boogert et al., 2013; Taylor et al., 2013), theory of mind (Andrews, 2018), and logical inference (Lazareva, 2012), to name a few.

In practice, these controversies tend to follow from at least two recognized issues in the literature. The first, which I'll label "the robustness problem," is the frequent tension that exists between studies purporting to test the same capacity. In a pattern that's been compared to the well-known "replication crisis" (McGetrick and Range, 2018), animals in cognitive studies will succeed in one version of a task only to fail when assessed using a slightly different design (e.g., setting a puzzle up vertically rather than horizontally; Taylor et al., 2012) or accomplish amazing feats in the field only to fail on simple problems in the lab (Suddendorf, 2013b, ch.7). Studies that seem like they should track the same mechanisms end up with different results. Without the ability to establish clear patterns of success or failure, however, researchers are frequently left at a loss. Whether the evidence favors one mechanism or another depends on which study one credits, and while such inconsistencies can sometimes be explained away, this isn't generally the case. More often, it's left indeterminate what caused a divergence—whether an unreported quirk in one design, a gap in the animals' understanding, etc. (see, e.g., Heyes, 2015)—and to what extent differences between tests matter for studying the mechanisms of interest (see below). Frequently, the most one can say is that the evidence is "mixed," a stubborn fact that at once undermines the literature's capacity to converge on any particular mechanism and shrouds novel results with the anticipation of experimental rejoinder.

The second major issue, which we'll call "the underdetermination problem," is that, despite experimenters' best efforts, results are invariably met with sufficiently many and sufficiently plausible alternative interpretations for members of the field to withhold judgment. Tests meant to show that non-human primates represent others' mental states, for example, are standardly criticized by skeptics for failing to rule out strategies based on contextual cues (Povinelli and Vonk, 2004; Heyes, 2017). Instead of taking others' perspectives or reasoning about beliefs, the subjects could simply be tracking patterns in their behavior (noting line of gaze or when they were last in an environment). Results cited by doubters, meanwhile, are commonly challenged for the artificiality of their designs (Call and Tomasello, 2008). In attempting to control for confounds, it's argued, skeptical researchers end up introducing them by testing subjects in wildly unnaturalistic and potentially confusing settings. The respective groups' criticisms are incorporated into each other's subsequent designs, of course, but results always seem to leave the door open for some confound or another. (Indeed, the debate has been so thorny that some have questioned whether it's an empirical dispute at all [Buckner, 2014; Halina, 2015]). Nor is this state of aporia exceptional, as a similar story could easily be told of research on self-awareness (Huttunen et al., 2017), metacognition (Smith et al., 2014), and a range of other areas where the dialectic of "romantics" and "killjoys" has become the norm (Balter, 2012).

Underdetermination and evidential conflict are not unique to animal cognition, of course. Both have been discussed in philosophy of science for some time. Examining the sorts of controversy usually encountered in the philosophical literature, however, one finds that they look rather different from animal cognition. For one, cases are frequently drawn from the cutting edge or most policy-relevant branches of a science, where friction can be attributed in part to the need to make recommendations on the basis of a time-restricted range of evidence. That disagreement often accompanies drug safety testing, for instance, is relatively unsurprising given that investigators are expected to form judgments about novel treatments in a matter of months (Biddle, 2013). The central disputes in animal cognition, by contrast, have been around for decades, with many addressing questions and methods developed in the sixties or seventies (see, e.g., Gallup, 1970; Premack and Woodruff, 1978). In the case of episodic memory, the late Endel Tulving once went so far as to suggest that "no essential progress has been made…for more than a hundred years" (2005, 5). Overstatement though it is, it's hard to even imagine such an idea being entertained in an area like experimental medicine. Looking to areas that do involve such long-lived controversies, moreover, one finds that the problem generally derives from distinct epistemological circumstances. There has been much discussion of controversy in historical sciences like paleontology and archeology, for instance (Currie, 2018), as well as sciences that study large, complex systems (e.g., seismology; Kleinhans et al., 2005). These may be traced, at least in part, to the difficulty of performing controlled and repeatable experiments on the objects of study, though. Without experimental control, it's harder to screen out environmental confounds or generate the precise circumstances needed to test a claim (Currie and Levy, 2019; Hacking, 1983, ch. 9-10). In the historical case, evidence is also subject to decay, allowing critical data to be corrupted or lost to time (Turner, 2005). Neither is much of an issue in the decidedly experimental science of comparative cognition.

Turning to the cognitive literature itself, the most common explanations are those implicating some aspect of cognitive theory or data analysis. In the first case, the discipline's problems are blamed on essentially Popperian failings. Critics have blamed the use of supposedly unfalsifiable notions like "understanding" (Penn, 2011) or, absent a specific formalism, "association" (Haselgrove, 2016), charging such intuitive appeals with muddying the theoretical waters. Others have pointed to the habit of shifting proposals post-hoc (Beran et al., 2016; Hanus, 2016) and the use of hunch rather than explicit derivation in the framing of experimental designs (Heyes, 2012), arguing that such loose practices render decisive results all but impossible. Were the competing ideas more clearly defined, all such proposals seem to suggest, researchers would be in a better position to find clear and stable experimental dissociations between proposals—to develop *experimentia crucis*. In concert with this push for a more predictive and computationally explicit comparative cognition, many have also turned to data analysis as a means of resolving the impasse. Some have argued that replication failures would be cut down by increasing the number of subjects used and instituting stricter standards for statistical significance (Stevens, 2017). More numbers, the reasoning goes, mean less noise and fewer conflicted results. On the underdetermination issue, some have also argued that model selection criteria like the Akaike Information Criterion (AIC) might help adjudicate between proposals (Sober, 1998), provided the theories can be formalized enough to utilize such tools.

Clearer theoretical framings and more sophisticated analyses couldn't hurt, of course, but if I am correct, it's doubtful these factors fully explain the issue. The specific reasons will become clear in later sections, but a few points may be noted at the outset. On the matter of replicability, we may note that animal cognition already has relatively good practices, with researchers routinely incorporating replications and extensions of others' work into their publications (McGetrick and

Range, 2018). Nevertheless, puzzling juxtapositions of results and disagreements over which versions of a test are legitimate persist (the many versions and results of the mirror self-recognition literature offering one example; Huttunen et al., 2017). Many of the conflicts, this suggests, result not from random noise but unaccounted for features of the mechanism or experimental set-up. While increasing the number and stringency of tests would likely cut down on the number of puzzling/contradictory results, then, it's unlikely they would remove the most long-lived problems. When it comes to theory, meanwhile, we know from prominent historical and philosophical investigations that the supposed shortcomings of informal theory and post-hoc accommodation are perfectly consistent with normal scientific consensus-building. Qualitative models are the norm in many *prima facie* successful areas of biology (Winther, 2006), while exploration, hunch-following, and post-hoc adjustment seem to facilitate scientific progress at least as well as strict hypothesis testing (Hacking, 1983, ch.12; Steinle, 1997; Franklin, 2005). It's unclear why animal cognition should be different. More strongly, and as we'll discuss more later, there is a case to be made that formal theory generally *follows* rather than precedes progress in the realm of experimentation (at least in biological systems; see Burian, 2007). Pushing for formalization as a means of effecting such progress may put the cart before the horse.

3.3 A Missing Factor

So, what's missing? One leading factor, I'd suggest, is measurement. As the principal channel through which researchers obtain evidence, measurement has long been recognized as a factor in scientific consensus-building. It's well-known that the development of classical thermodynamics was enmeshed with the construction and standardization of thermometers

(Chang, 2004, ch. 4; van Fraassen, 2008, ch. 5-7) and that much of modern medicine and microbiology depended on the gradual refinement of microscopes and staining techniques (Pannese, 1999). But when it comes to *controversy*, explicit and sustained discussions are comparatively rare (Miyake, 2013).[39] The dramatic conflict of ideas still gets more attention than the concrete channels through which scientists probe their targets. If I am correct, the latter has as much of a role to play in the creation and endurance of controversies as the latter. To see why, though, we'll need some background on the epistemology of measurement.

3.3.1 Dimensions of Measurement

Abstractly, one can think of measurement as an interaction between a target system, an instrument, and environmental conditions that produce a set of readings (a pointer on a gauge, an array of numbers stored in a computer, etc.). Under certain assumptions, these readings may be used to construct a representation of the target features (e.g., an image; Tal, 2015; Frigiero et al., 2010) and provide the researcher with some useful information about the system studied. A micrograph derived from a sample of liver tissue, for instance, will convey information about its cellular structure under the assumption that the preparation doesn't damage the tissue, that its spatial properties are not distorted by the instrument, and so forth. Whether a measurement process is adequate for one's epistemic purpose will, in turn, depend on both the nature of the information

---

[39] Following Miyake's work, measurement has been cited as relevant in recent reviews of uncertainty in geoscience (Bokulich and Oreskes, 2017). Relevant discussions may also be found in the classic philosophy of experiment literature (e.g., Hacking, 1983; see below).

conveyed and the justification that experimenters can give for their assumptions. Speaking more precisely, adequate measurement depends on meeting at least three standards.

The first two benchmarks concern the information researchers can access. Though terminology differs somewhat by field, it's standard in many sciences to assess instruments with respect to their sensitivity and selectivity. As used here, the former indicates the degree to which differences in the target variable(s) are reflected by differences in the instrument's readings, while the latter indicates how *insensitive* the tool is to non-target or exogenous variables. Roughly, the difference is between picking up relevant features and screening out interference. In the case of sensitivity, a further distinction may be drawn between low and high-dimensional varieties (Cf. Franklin, 2005), the first indicating tools sensitive to some *specific* aspect of the target (e.g., thermometers) and the second indicating those that simultaneously register *many* aspects (e.g., microscopes). The less sensitive an apparatus and the lower its dimensionality relative to the target process, the lower one's capacity to describe or manipulate the object of study. Without a high enough resolution microscope, for instance, one wouldn't be able to describe the morphology of a cell or perform experiments that require injecting foreign genetic material into its nucleus. Such limitations naturally place one's research program at a greater risk of underdetermination problems, as the descriptions and manipulations rendered unavailable may be just the evidence needed to push an idea into widespread acceptance. They could also generate robustness issues, as the less sensitive one's instrument is, the greater one's chance of missing target features or changes relevant to the outcome of an experimental intervention. Two labs might find radically differing results if, say, they'd inadvertently acted on different sites due to a lack of microscopic sensitivity. The same can happen for selectivity, as the lower the instrument's ability to screen out interference, the greater one's chances of receiving mixed signals. Interference in a radio signal, to take an obvious example, can lead people tuning in at different times and places to hear markedly different broadcasts.

Turning to validation, we face a very different set of questions. Instead of looking “from above” at the instrument and what we believe it’s capable of conveying, we look at the situation “from within,” examining the justification researchers in a given historical or disciplinary context can give for trusting it (van Fraassen, 2008, p. 139). An instrument might be sensitive and selective enough from the perspective of a god, but if the researchers have no reason to think it is, it’s unlikely they’ll put much faith in its readings. A result, one might say, is only as strong as the justification available for the process used to obtain it (Miyake, 2013). What justification is available depends on a variety of contextual factors. Prominent examples include the convergence of independent measures, the successful use of the tool in striking interventions, and the appeal to some well-developed theoretical background for the measurement procedure. The buildup of trust in thermometry, for instance, was backed by the fact that multiple thermometers—including those composed of differing substances—could be checked against one another on the same target at the same site (Chang, 2004, ch.2), while light microscopy, which lacked independent measures for many years, could be supported by experimenters’ ability to intervene on the targets and the conditions surrounding them—poking, prodding, shifting the lighting used, and so forth (Hacking, 1983, ch.11). In still other cases, tools have been justified by reference to the theories that underwrite their operation or corroborate their results. The development of radio telescopes, for example, was backed both by the engineering that went into their design and the fact that their results cohered with strong theoretical arguments regarding the phenomena of study (e.g., about Jupiter’s radiation patterns; see Franklin, 1989).

No one validation procedure appears strictly necessary, but without the support of either a multiplicity of checks or an exceptionally strong backing in one or another domain, measurement techniques encounter at least two problems. First, they're subject to doubt and interpretive controversy. Without some evidence that the instrument is tracking what users claim it does, critics are free to offer alternative interpretations of its results. The existence of the Golgi apparatus near the nucleus of the cell, for example, was subject to doubts for more than fifty years for this reason. Though present in certain stains, skeptics worried that the organelle's appearance might have been an artifact until electron microscopy was able to provide independent confirmation (Farquhar and Palade, 1981). Second, procedures that haven't been validated will be liable to inconsistent or inappropriate application. Every instrument has limitations—levels of temperature, size, or setting beyond which their readings cease to track their targets. Validation is what allows us to sift these uses out, and without it, labs operating in different contexts are liable to encounter mixed or inconsistent results. The early history of microscopy, for instance, reveals many a dispute bred of what, in retrospect, appear to be unrecognized artifacts, a lack of standardization, and the misapplication of tools (mistaking lens aberrations for worms, aiming to study the structure of light with optical microscopes, and so on; Wilson, 1995, ch. 2, ch. 7; van Fraassen, 2008, ch.4). It's only *after* the development of validation procedures like those listed above that one sees these problems start to abate and the community of users reach consensus on their legitimacy and proper use.

Summarizing the foregoing discussion, we have at least three non-exclusive and non-exhaustive ways that measurement complications might sustain or engender conflict:

I. The instrument may be insufficiently sensitive, yielding ambiguous results and/or failing to detect relevant changes in the target.

II. It may be insufficiently selective, introducing confounds and causing inconsistencies in the data.

III. The researchers using the instrument may be unable to validate it and unable to build confidence in their claims because:

    i. They have yet to take concurrent, independent measures to check its readings

    ii. Their preferred tool limits their ability to perform and trace the impacts of interventions

    iii. They lack knowledge about the tool's operation and/or the target system that could be used to check its readings

That such features could lead to scientific conflict is, I suspect, not terribly controversial. That fuzzy images or mixed signals leave room for disagreement is familiar enough from everyday life. Whether these factors are major players in animal cognition isn't so clear, though. To establish this, we need to know more about the relevant tools and their uses.

3.3.2 Dimensions Applied to Animal Cognition

As was mentioned earlier, animal cognition represents an interesting mix compared to related traditions like classical ethology or behavior analysis. Target-wise, its aims present a marked contrast with its neighbors. Most other areas can be described as studying behavior directly. When researchers sift through communication patterns or document activities like tool use, say, they are studying targets that can ultimately be defined as sets of actions. Cognitive processes, by contrast, are traditionally understood as *causing* rather than being constituted by behavior (Leavens et al., 2019): the targets aren't the actions themselves, but the proximal, psychological mechanisms behind them (Shettleworth, 2001). In terms of its research methods, however, the field looks a lot like other areas, with most investigations drawing on some combination of field observation and behavioral tests run in labs and zoos. Rather than probe cognitive architectures "directly" via neural recording or lesions, researchers infer their structure and operation through their downstream behavioral impacts. In effect, the animal's behavior serves as a "reading," its motor system as the "instrument."

This might not seem like a terrible restriction at first. Behavior certainly does reflect thought, and as humans, the move from cognition to action can seem almost transparent. We possess a highly refined motor system featuring, among other things, a uniquely large number of direct corticomotoneuronal connections for fine-grained control of the face, tongue, larynx, pharynx, and hands. For most non-human vertebrates, however, the bulk of action selection relies on less direct and evolutionarily older pathways through the basal ganglia and other subcortical structures (Holstege and Subramanian, 2016).[40][41] These mechanisms tend to be relatively small, receive inputs from numerous sources, and produce a relatively limited motor repertoire (Oorschot, 1996; Zheng and Wilson, 2002). Input-wise, signals come from across the cortex as well as from sub-cortical structures like the amygdala, yet in terms of motoric outputs, they all pass through the same handful of nuclei in the brainstem and spinal cord. Ultimately, this series of many-to-one mappings ensures that signals from distinct upstream architectures can shift and combine while

---

[40] Other primates share some of our specialized hand and upper-limb circuitry, but the difference in quantity is considerable (Lemon, 2008).

[41] Many of drawbacks reviewed here are also present in human behavioral research and may account for some controversies there (Anderson, 1978; Boone and Piccinini, 2016). The issues appear to be mitigated by the anatomical refinements and our ability to communicate with subjects.

still producing smooth, consistent movements (Redgrave et al., 2010). Evolutionarily this makes sense—one wants transitions between decision-making mechanisms to be relatively free of breaks and instabilities—but from the perspective of measurement it has considerable drawbacks.

In the first place, the extreme winnowing that accompanies the move from thought to action imposes severe limits on the amount and dimensions of cognitive variation to which behavioral measurements are sensitive. There is only so much room, and as more information is lost, one's risk of underdetermination increases. In terms of which dimensions are registered, we may note that, trivially, the motor system favors those aspects most immediately relevant for guiding and executing movement—the target of a reach, the urgency of a movement, and so forth. *Prima facie*, however, there are many important aspects of cognition that don't fall into this category—differences in the division of labor between upstream architectures, say, or what mechanisms coordinate these operations—and if these dimensions are the sole differentia between architectures, it follows, behavioral evidence will underdetermine one's choice between them. There is a very real possibility of distinct yet behaviorally equivalent architectures. Even if there were clear behavioral differentia for cognitive processes, moreover, the amount of information conveyed by behavior could still be the too meagre to ground compelling inferences. Consider the case of dreaming. The comparatively "high-throughput" channel of the human motor system allow for compelling behavioral evidence of dreaming in humans. In most cases, one need only ask. Given their more limited motor repertoire, however, the principal behavioral evidence for non-human dreaming comes by way of rapid eye movement. Now, for all we know, REM could be a highly reliable indicator of dreaming vs non-dreaming sleep. The human data would seem to suggest a close relationship. The behavior is hardly a smoking gun in the case of other animals, though, as the information conveyed about upstream processes is so minimal (Suddendorf, 2013a,

p. 45). The systems simply aren’t made to communicate trains of thought in the way that human anatomy seems to be, and given this, it’s not surprising that the field should have more of an underdetermination problem. The likelihood of evidential ambiguity given the targets studied and tools used is fundamentally higher than either human cognition or narrowly defined animal behavior research.

A similar story can be told for robustness. Echoing the discussion in 3.2, we may note first that conflicting results may stem from overlooked features or changes in the target’s behavior. The precise structure and operation of the mechanisms being studied are still largely unknown, so it’s difficult to anticipate how changes in task or environment will impact them. With temporally sensitive and high-dimensional tools, this need not be a problem. One can catch when or which changes occur and pin down the environmental differences eliciting them. Indeed, it’s quite common for scientists using high-dimensional tools to go about theory construction in just this way—tweaking parameters and seeing what happens instead of forming and testing explicit hypotheses (Franklin, 2005). With a less sensitive device, however, these interactions become much harder to identify and control for. Obvious confounds can be screened out, of course, but others will slip by. Without a sensitive screening procedure, for instance, one can end up lumping together subjects whose cognitive architectures differ before the experiment begins or who diverge midway through a trial, yielding mixed results for what seem like identical studies.

Examining the system’s selectivity, things only get worse. As I noted earlier, action selection is shaped by a multiplicity of computationally and anatomically distinct mechanisms (Van der Meer et al., 2012). It’s not obvious from the outside which is shaping behavior at any given time, however, as their signals are quite literally mixed at the level of motor control. This, combined with the fact that we don’t know precisely how changes in an animal’s environment influence the switch from one system to another, means that studies which look very similar could end up with very different results. Differences of temperament, history, and hard-to-anticipate environmental properties may all cue distinct combinations of guiding systems (Packard, 2009; Packard and Goodman, 2013), yielding an evidence base that's mixed not because of anything to do with the target but because the measurement techniques allow interference. As with the underdetermination problem, the robustness issue is something that would be expected were measurement the driving issue and is consistent with what we know about the instrument.

Finally, turning to validation, we find that the field's focus on behavioral over mixed methods leaves it vulnerable. We can see this by considering the typical routes by which measurements are checked: theoretical backing, simultaneous measurement, and intervention/interaction. As was just emphasized, there's still a lot we can't predict about the motor system's operation, so a theoretical justification of the kind seen in the physical sciences is out of the question. Nor are standard intervention and parallel measurement-based checks available, as both tool and target are embedded in the organism. One can't assess two motor systems on the same target or, short of invasive techniques, check behavior against fine-grained manipulation or independent measurements. This leaves the purely behavioral researcher with relatively few resources, and in practice, researchers end up relying on ecologically informed judgments of face validity (i.e., whether the behavior seems like it should track the target). This doubtlessly rules out some inappropriate designs, but for the most part, the vulnerabilities listed earlier remain unguarded. We can recognize that throwing an animal into a pool of water is unlikely to encourage reflection, but opinions will differ on whether long-term practice using mirrors may reasonably be included before testing self-recognition (Chang et al., 2017), whether evidence of foresight requires single-trial studies (Scarf et al., 2014), or how seriously to take the ubiquity of cages and

Plexiglas barriers in cognitive studies (Leavens et al., 2019). In terms of knowing when and how to use the tools, the situation is not unlike the early microscopists'.

3.4 Case Study: The Cognitive Map

Reviewing the evidence above, it would appear that the anatomy of many non-human subjects' motor systems have substantial shortcomings in sensitivity and selectivity, that their structure and embeddedness in the organism make them difficult to validate using purely behavioral means, and that both underdetermination and robustness problems are consistent with these limitations. Hypothetically, a system like that used in animal cognition would be expected to engender and sustain controversy. To cement the case, however, we'll want more than hypotheticals. What's needed is evidence not simply that measurement problems could generate controversies like those of contemporary animal cognition, but that it has. To this end, one can find no better case than Edward Tolman and his famous cognitive maps. Tolman's work is, in many ways, the prototype for cognitive research. In an era dominated by purely behavioral accounts, he self-consciously advanced a program focused on internal processes and structures that would come to define the cognitive movement (Amundson, 1983). At the same time—and in contrast to some of the still-ongoing debates listed earlier—his proposal that humans and other animals use structured representations in problems ranging from navigation to social decision-making is now widely recognized as accurate (Eichenbaum et al., 2016; Schiller et al., 2014). Indeed, the characterization of Tolman's "cognitive maps" earned several psychologists a Nobel Prize in 2014. "All three," as honoree Edvard Moser noted in his memorial lecture, "stand on the shoulders of E.C. Tolman" (2014, p. 402). His ideas weren't always so celebrated, though.

Behavioral researchers voiced doubts about his approach throughout his lifetime and his ideas continued to be controversial well into the present century (Shettleworth, 2010). Understanding how this situation came about and how it was resolved thus offers an ideal test for the claims laid out above. I'll start with a brief overview of the controversy, returning to measurement after.

3.4.1 Outlines of the Controversy

At base, Tolman's proposal was simple. Rather than (or in addition to) learning a collection of stimulus-response relations, he argued, animals like rats tend to solve problems by using topologically structured mental representations. These architectures, which he likened at different points to a perceptual gestalt or "a field map…in the rat's brain" (1948, p. 192), would integrate task-relevant knowledge or "hypotheses" into a broader network. Exploring a maze, for instance, would allow a rat to form a map that encoded spatial relations, while learning to press a lever for food might encode simple cause-effect relations (1959, p. 132). Instead of being bound to specific motor sequences, these maps could guide a variety of actions depending on the animal's goals (exploration, searching, etc.). Likewise, instead of having to relearn an entire system of responses in the face of environmental change—the introduction of a barrier, say—they could be selectively revised by updating the inaccurate "hypothesis" (e.g., that hallway X connects with room Y). He even suggested that his maps might allow a kind of forward-directed thought. Faced with a choice of which path to take, rats might use the map to perform what Tolman, following colleagues Muenzinger and Gentry, labelled "vicarious trial and error" (VTE)—a kind of mental simulation of the available alternatives (Tolman, 1948).

By all accounts, the suggestions were radical for their time. Many of his contemporaries were warry of *any* talk of internal, mental mechanisms. As a trained behaviorist, however, Tolman was sure to lay out clear behavioral criteria or "defining experiments" for each claim (Tolman, 1948)—a practice shared by subsequent cognitive researchers (Shettleworth, 2001). Patterns of back and forth head movements were used to show VTE at decision points, "searching" for missing objects to make the case for hypothesis updating, and navigation in situations requiring novel behaviors to demonstrate that rats learned representations of "place" rather than specific behavioral dispositions (see Tolman, 1948; 1959). Most famously, he and dozens of later psychologists sought to establish the existence of environmental maps by studying animals' use of detours and shortcuts, the thought being that these capacities would exceed what alternative mechanisms were capable of (Morris, 1981). From today's perspective, Tolman's claims were largely on-point. There are, in fact, topologically organized structures in rats' brains (the hippocampus, specifically) that map onto locations and/or features in their environments. Each "map" is composed of cells responsive to a location (object, conspecific, etc.) that, through their interactions, encode relations between these variables in a functionally integrated network. These maps can now be imaged in some detail, and when assessed in contexts cited by Tolman, they've tracked his predictions with an often-startling accuracy. Rats "searching" for missing objects do manifest heightened activity in corresponding cells, which recedes as they habituate to the objects' absence (O'Keefe and Nadel, 1978); novel barriers and pathways do involve adjustments to the maps' topology that can be exploited for things like shortcut discovery (Alvernhe et al., 2008; 2011; Pfeifer and Foster, 2013); and remarkably, back and forth head movements do correlate with something like mental simulation, as cellular firing during such episodes ceases to track the animal's current location and traces its way, one by one, down the paths ahead (Redish, 2016).

This underlying accuracy proved to be of little help, however, as Tolman and his followers were embroiled in one of the defining controversies of the era. During the psychologist's lifetime, his proposals were routinely dismissed (Restle, 1957), and while their standing did improve after the "cognitive revolution" of the sixties and seventies, their existence was never close to settled in the behavioral literature. Well into the twenty-first century, one could find serious challenges to cognitive mapping, with respected scientists arguing that it was under-supported and suggesting that progress would be better served by abandoning the notion in favor of something better-specified (Cheng et al., 2007; Shettleworth, 2010). As one might expect from today's controversies, the reason psychologists were unable to agree came down to two key issues: robustness and underdetermination.

Like so many contemporary disputes, critical studies routinely failed to replicate or generalize. Conducting studies using a slightly different maze (Gentry et al. 1947), changing the number of external landmarks available (Benhamou, 1996), or switching the animal's motivations (Hill and Thune, 1952) could dramatically alter a result. There were also a variety of studies that, while not strict replication failures, conflicted strongly with Tolman's vision of the maps' operation. While Tolman's rats seemed to flexibly adapt to changing environments, those of Carr and Watson (1908) would run headlong into a wall if their habitual path was shortened; while the animals studied at Berkeley were clever enough to transfer their knowledge of a maze's layout from running to swimming, those in Iowa were so clueless that they couldn't recall the location of food they'd seen while thirsty (Kendler, 1947). To some extent, this variability was found even *within* labs, with predictions about things like the timing of VTE conspicuously failing to generalize across set-ups (Tolman, 1948). Without a consistent pattern of evidence, defenders and critics both had precious little on which they could build a compelling case, leaving doubt, frustration, and the controversy that invariably results from their mixture.

At the same time, the results on which Tolman built his case usually had alternative interpretations. Looking back and forth, to take the most striking example, is consistent with the rat engaging in a process of deliberation but could just as easily reflect simple indecision. For many, the very idea of grounding such a grand inference on so slim an evidential base seemed laughable (Guthrie, 1935, p.172). Likewise, rats *could* be consulting some internal map in taking shortcuts, but they might just as easily have learned responses to various internal and external cues: relying on visible landmarks to indicate which way to turn, using a process of "dead reckoning" in which kinesthetic and/or vestibular information guide responses, or employing some combination of factors (Bennett, 1996). Psychologists tried to pull the proposals apart for decades but kept encountering the same problem: to use a map, one needed landmarks or self-motion cues to stay oriented, but so long as these cues were available, one could argue that the animal was using them more "directly" as guides for action. Just as with contemporary theory of mind research, the competing architectures were simply too similar in the inputs they needed to operate and the outputs they could provide. Some even suggested that the problem might be semantic rather than empirical (Restle, 1957).

3.4.2 Explaining the Controversy

During Tolman's lifetime and for decades afterward, these problems persisted. Waves of studies came and went, but the reason for the literature's robustness and underdetermination problems remained a mystery. For some, the problem was that Tolman's suggestions were too vague, his task designs far too intuitive, and his standards of evidence too low (see, e.g., Restle, 1957; Bennett, 1996; Cheng, 2007). As late as the present decade, in fact, prominent figures in the behavioral literature suggested that cognitive maps might be an anthropomorphic holdover, something whose track record might even recommend abandonment (Shettleworth, 2010). In retrospect, however, such judgments are almost certainly mistaken. Better theories and higher standards of evidence may not have hurt, but as I'll argue shortly, it's unlikely they could have settled the matter. This is because, at a fundamental level, the problem was with the instruments. For all the reasons listed above, behavior failed to provide robust and unambiguous evidence of cognitive maps because the anatomical systems on which it relied were insufficiently sensitive, under-selective, and without clear validation. Without changes to these, Tolman—and those using the same tools for similar ends—couldn't help but run into robustness and underdetermination issues.

We can start with sensitivity. As was noted above, the motor systems of most animals have limitations when it comes to the information they convey about upstream cognitive processes. Rats are no exception, and examining the history, it's easy to see how this negatively impacted Tolman's research. Most straightforwardly, the lack of relevant information led to evidential ambiguities. Measured with sensitive, high-dimensional instruments like the multichannel electrodes used in contemporary neurophysiology, phenomena like VTE and shortcut discovery are quite striking. One can effectively "see" as activity in the hippocampal maps moves down alternate paths or puts together a new sequence of cells to integrate previously unconnected regions (Redish, 2016; Pfeiffer and Foster, 2013). Little of this makes it out of the hippocampal formation, however, as the region has no direct impacts on behavioral control (relying on the ventral striatum and prefrontal cortex as mediators; Goodroe et al., 2018) and the sharp-wave and low-theta oscillations that underwrite both processes are generally limited to periods of immobility (Lever et al., 2014). The real action takes place in dimensions of the cognitive process that motor control simply isn't privy to. Wavering head movements excepted, it's not until after the process has terminated (i.e., after the rat has decided) that behavioral manifestations emerge, and given the information bottleneck imposed by the rats' motor system, these prove quite limited. The substantial array of information relayed from the hippocampus to the ventral striatum ultimately boils down to a handful of actions (e.g., turning left instead of right). Perhaps there is some as-yet undiscovered way to characterize cognitive maps using such sparse data, but at the very least, cognitive researchers' decades-long focus on behavioral designs over mixed methods made the controversy far more difficult to settle.

Less obviously, the motor system's insensitivity also contributed to the robustness issue. As we now know, maps can be quite volatile. If the environment undergoes some but not other changes or if the animals switch between tasks that are just a bit too dissimilar, the maps may undergo an instantaneous "remapping" in which cells come in part or in full to represent new locations or features. Functionally, this is thought to help prevent interference between memories needed for different tasks or relevant to distinct locations (Colgin et al., 2008), but it also means that useful information won't always be recalled. The fact that remapping occurs with certain landmark changes (Bostock, 1991), for instance, means that rats may have difficulty locating a goal after major landmark adjustments (a robustness failure reported by Benhammou, [1996]). Likewise, the fact that remapping occurs between shifts in motivation, including between water- to food-searching (Kennedy and Schapiro, 2009), likely accounts for rats' failure to locate food sites visited while thirsty (Kendler, 1947). Were there some outward indication that the cognitive target had undergone such massive changes, these apparent robustness failures might have been less problematic. As it was, however, it was hard not to regard the behavioral inconsistencies as a serious challenge. To insist that some unevidenced remapping had occurred would have smacked of special pleading.

The issue was only exacerbated by the system's limited selectivity. Hippocampal maps are one mechanism influencing decision-making, but as I mentioned earlier, there are several others, including Pavlovian conditioning mechanisms in the amygdala and a habit-like learning system in the dorsolateral striatum (Van der Meer et al., 2012). The differences between these systems is hard to miss when looking at the recording data or the effects of pharmacological intervention. From the outside, however, it's not clear which behaviors share an upstream cause and which don't—or even how many upstream mechanisms there are. It isn't as though the behaviors guided by dorsolateral striatum all share some tic. Without a natural means of parsing the behaviors by cause, however, different labs ended up studying different targets. Reviewing the literature, one finds that a large portion of its robustness problems were traceable to this fact. Curious, seemingly map-inconsistent behaviors like running into walls are now seen as a product of mechanisms altogether separate from those Tolman sought to explain (Smith and Graybiel, 2013), as are many of the would-be replication failures. Differences in training duration, environmental cues, and various other tweaks that served to reverse results in the behavioral literature have all been found to shift the balance of hippocampal and dorsolateral striatal influence. In the circumstances used by Tolman's critics, the striatal, habit-related system is upregulated, while in circumstances closer to his supporters' set-ups, the hippocampus gets the edge (see Packard and Goodman, 2013). The results were mixed because the targets were.

Strikingly, this was something that many in the literature suspected. Authors like Schnierla (1952) suggested early on that different labs might be studying different mechanisms, and Tolman himself argued that conflicting findings stemmed from different maps being "induced" by shifts in motivation or training (Tolman, 1948, p. 205-7). The problem, as Tolman observed at the end of his career, was that experimenters had no way of ensuring that different studies captured the same processes: one simply couldn't predict how the relevant variables' interactions would change from one set-up to another (1959, p. 148). In later years, authors in the physiological literature would couple their behavioral observations with recordings, lesions, and pharmacological intervention (O'Keefe and Nadel, 1978), ensuring that each study had a common target by "sacrificing" the animal and staining the recorded or lesioned tissue. For Tolman and subsequent authors in the behavioral literature, however, there was little one could appeal to in connecting some results and setting others aside. Provided the resulting architecture wasn't *too* Baroque, researchers were free to carve-up the behavioral data how they liked, attributing one finding to one mechanism and conflicting results to another. While not totally unconstrained, the absence of independent measures left more than enough room for reasonable people to disagree.

Given all of this, I'd argue, it's unlikely the controversy could have been quelled by anything other than a shift in instrumentation. Larger sample sizes and stricter standards of evidence might've helped eliminate some of the variance in results, but the most pressing behavioral inconsistencies seem to stem from genuine features of the system rather than chance. Rats switch between maps as well as between map and non-map mechanisms, and this was going to produce different behavioral patterns regardless of how one analyzed the data. The effect of more rats and lower significance thresholds would simply have been to entrench the conflicts. So too with theory. Tolman has long been criticized for his imprecision, but even if he had developed a sophisticated account—an account such as we have today—he probably couldn't have convinced his peers. The myriad biological assumptions and parameters that go into present-day theories of mapping would appear far too speculative without the anatomical and physiological data that back them up (something that would hold both for qualitative assessment and formal assessment via measures like the AIC). What's more, there's only so much he could have predicted given the motor system's limited sensitivity. Tolman could have provided elegant explanations of things like shortcut discovery, detour-taking, and long-distance navigation, but these behaviors can all be accounted for by alternative mechanisms (Cruse and Wehrner, 2011). Underdetermination and mixed results were practically unavoidable so long as researchers continued to rely on behavior. Qua instrument, the rat's motor system was and is too insensitive to cognitive variables, too prone to mixed signals, and too difficult to validate without non-behavioral tools or interventions. Insofar as today's cognitive researchers are using the similar tools for similar ends, it's unsurprising that they too face underdetermination and robustness problems.[42]

3.5 The Cognitive Dilemma

The field of comparative cognition was born in deep controversy and, despite its ascendance, never seems to have escaped it. The debates over cognitive architectures and processes that it ushered (or reintroduced) into comparative psychology have been the site of exceptional creativity over the years—including some of the most striking experimental designs in psychological science—yet they have not, by and large, led to closure or robust scientific consensus. Basic questions, some of which were with the field at its birth, remain the topic of intense disagreement, and major publications routinely express frustration at the feeling of too little progress. The reasons for this have been debated for nearly as long as the questions have,

---

[42] Indeed, many processes surrounded by controversy—imagination, logical reasoning, causal reasoning, and various forms of social cognition—seem to depend on the very same circuits as cognitive mapping (Rubin et al., 2014).

though discussion has traditionally focused on a common set of culprits and recommendations. The failing, it’s long been argued, lies somewhere in the realm of ideas: in the articulation of the theories tested, the derivation of predictions, or the assessment of data. Were the field to insist on greater rigor in these domains, it’s said, the controversies that surround the field may abate. If I am correct, the issue is more material and, in a sense, lies closer to the cognitive paradigm itself.

The hope of the approach, from Tolman through today, was that the behavioral methods could be used to address questions not only about behavior but about the proximal, mental mechanisms that cause it. What the physiological and historical evidence reviewed above suggests is that this isn’t so. Assessing the mechanisms governing behavior against standards common in the physical and life sciences, I’ve argued, one finds stark limitations, including considerable limits to sensitivity and selectivity and a relative paucity of validation procedures. Each shortcoming, I’ve argued, increases the odds that a study or body of research will face ambiguous results, evidential conflict, and simple doubt. In examining the paradigmatic controversy over cognitive maps, moreover, the combination of all three was found adequate to generate a pattern of robustness and underdetermination problems qualitatively identical to that of today. The way the cognitive map controversy was resolved (and in retrospect, the only feasible way it could have been) was through sustained attention to the development and validation of measurement techniques. Sound behavioral design may be necessary for progress (as argued by Krakauer et al., 2017) but likely won’t suffice.

In and of itself, this fact does not imply that students of animal cognition must take up non-behavioral tools. Such a choice can only be made in light of further, distinctively moral considerations about the use of such techniques.[43] It does, however, imply that the relative insulation of cognitive theorizing from neuroethological evidence seriously undermines the field's aim of characterizing the mind's inner workings. The cognitive programme's historic goals may be asking more than its historic tools can provide. Insofar as we think it fair to blame individuals only for outcomes under their control, moreover, the arguments above suggest that the tendency to fault theorists for their lack of precision or inability to derive *experimentia crucis* is misplaced. With a more favorable set of measurement techniques, the very same folk notions and intuitive tasks could guide a great deal of experimental progress. With better tools, theorists would have more to work with: something that could narrow the field of possibilities enough that constructing precise computational models wouldn't involve such outright speculation or risk of misspent effort. The decision may be made to forego neuroethological methods of the kind discussed above, but if so, the field should be less critical of theoretical shortcomings and more accepting of uncertainty.

---

[43] Even if adoption of invasive techniques is ultimately decided against, existing data about the structure and operation of different animals' neurocognitive architectures as well as any behavioral correlates they have may be used to constrain the space of viable theories (Smith and Church, 2017; Hasz and Redish, 2018).

4.0 Fourth Paradigm: Building a Neurobiological Case for Animal Rationality

4.1 Introduction: Dog Logicke

Buried in the history of Queens College is an episode curiously familiar to those working on animal minds. In March of 1615 King James I is said to have attended a Cambridge disputation concerning "the discourse and logicke of the dogs." It was a topic "tempered and fitted to his content," and with an eye to this affinity, speakers were chosen from the "ablest in every faculty." In support, a young doctor Preston recited what by then was a well-known case. Following the ancient stoic Chrysippus, he noted that a hound on the trail of a hare could be seen to engage in active deliberation. Having the major premise in mind that the hare must have gone down one or another path, it proceeded to consider each option, filling the minor by means of its nose. By eliminating potential paths, the thoughtful hound realized that hare could have gone but one way and bolted down the path with open mouth. The Answerer, one Dr. Wren, remained unimpressed. The dog may *appear* to reason, he allowed, but success in the chase presented "sagacity" more than "sapience." Hounds, he noted, had been endowed with acute senses and natural wisdom in matters concerning their stomachs, but beyond this, he argued, they had little to show. The debate, we are told, fast descended into such "wrangling about their Syllogismes" as to "put the King's acumen into streights." In the end, there was no clear resolution, though we are told the regent suspected from personal experience that "a hound had more in him than was imagined."[44]

---

[44] The full report is reproduced in Mayor (1898).

Despite the undeniable advances in the centuries since the storied debate, the discourse surrounding animals remains much the same today. There are a variety of behaviors that, on their face, suggest deliberation (even the sort of canine exclusive inferences cited by doctor Preston; see Aust et al., 2008; Erdohegyi et al., 2007). Yet as years of our own syllogism wrangling has shown, such observations are invariably subject to doubt (see chapter 3). What looks like logical inference from the outside can often be accounted for by some other mechanism. The animal might be guided by reinforcement (Vasconcelos, 2008), say, or an appreciation of subtle perceptual cues (Penn et al., 2008). In any event, it's argued, nothing so grand as deliberate logical inference need be posited. Objections meets reply, which in turn is met by further objection, producing a literature intricate enough to put even the most careful observer's acumen into straits. As vast and long lived as the debates have been, however, they have been unable to eliminate the suspicion of lay and professional audiences alike that something has been missed—that our own cats and dogs have more than the Wrens imagine.

The purpose of the present chapter is to vindicate these suspicions. If I am correct, a wide variety of non-human animals are rational creatures in the basic sense that they draw explicit inferences in deciding between competing lines of action. Human reason doubtlessly encompasses more than its animal counterparts (something I will return to in the postscript), but Darwin was fundamentally right in thinking the two were of a kind. Drawing on recent work on human reasoning, I break this deliberative process into two rough components, one executive and the other imaginative. The first is concerned with framing problems and the second with working them out. I then add two conditions drawn specifically from authors skeptical of animal reason—the capacity to employ abstract concepts and to draw inferences in a flexible manner. I then consider how these processes are biologically realized in humans, focusing on the prefrontal cortex and hippocampus, arguing that the same mechanisms are present in many other animals. Examining evidence from neural recordings conducted in the homologues of these structures in non-human animals, I argue that clear illustrations may be found of both the framing and imaginative components of reason and that these show precisely those characteristics designated "uniquely human" by skeptics. In the end, I conclude that many animals—perhaps even dogs—are reasoning creatures.

4.2 Reason and its Evidence

4.2.1 Isolating the Phenomenon

When Dr. Wren contended that hounds had "little in their myndes, unless it had relation to their mouths," he voiced a sentiment not uncommon at the time and not without power today. While the idea that animals are mere automata or "thoughtless brutes" has long been considered extreme (see Fudge, 2005, ch. 3), the view that something critical separates us from our evolutionary kin has not. Animals, as skeptics have traditionally allowed, benefit from acute perception and good memories for particulars and associations related to their present environs. When it comes to the "higher" faculties on which we pride ourselves, however—the capacity to think abstract thoughts and imagine future or distant states of affairs—humans and animals part ways. "Man above all other creatures," writes Wren's contemporary Grotius, "has been endued with judgement to discern things pleasant and hurtful…not only present but future and such as may prove to be *in their consequences*" (*De Iure Belli*, Preliminary Discourse, Para. IX, emphasis mine); "Beasts" Locke tells us, "compare not their ideas further than…sensible circumstances annexed to the objects themselves;" they "abstract not" and may fail to notice even the absence of a pup if perceptible signs like whining are not immediately available to them (*Essay Concerning Human Understanding*, Bk.II, Ch.XI).

Despite the centuries that separate us, many of today's discussions are not far removed from this picture. Our language has changed and the methods by which we assess claims of animal reason have shifted markedly, but for skeptics and the apologists who address them, the break between sapient, reflective humans and sagacious, perceptive animals remains at the literature's core. Animal cognition is commonly described as "automatic," "associative," and "experience-based," while humans are said to be capable of thought that is "controlled," "rule-based," "serial," and "consequential" (i.e., inferential; Evans and Stanovich, 2013). We alone have open-ended and "domain-general" processing instead of special-purpose adaptations (Dennett, 1996; Premack, 2010) or, as Grotius observed, the means to anticipate future possibilities (Suddendorf and Corballis, 2007). Humans, it's said, reason using schematic and flexibly applied cognitive structures, while animals think only in perceptually grounded or context-specific terms (Dennett, 1996; Halford et al., 2010; Millikan, 2004, ch. 1, ch. 15-18; Penn et al., 2008; Premack, 2007). It is precisely this sentiment and associated set of psychological claims that I mean to counter.

The sense of "rationality" at issue here is thus a descriptive one, dealing primarily with the mechanisms and processes of decision-making rather than their results or the standards to which they should be held. When humans are faced with tasks like planning a trip or selecting a move in chess, we engage in what Hurley and Nudds aptly characterize as a process of "decentring from the me-here-now and entertaining alternative possibilities" (2006, p. 3; see and Cf. Johnson-Laird, 2010, p. 18243; Simon, 1956, ch.3; Stanovich, West, and Toplak, 2014, p. 83). My claim is that other species do so as well. To make the case, I focus on two capacities that, when combined and subjected to qualifications outlined momentarily, yield just the sort of process one encounters in standard texts on reasoning. The characterization will not be exhaustive—frankly, the current state of the research renders a definitive account of reasoning premature—but seeing as our aim is to *identify* rather than *explain* the reasoning process, nothing so detailed is necessary. One needn't know just how vision works to know that cats see or have all the mechanisms of digestion charted out to recognize it in dogs; all that's required are features sufficient to differentiate the process from relevant alternatives. In the case of reasoning, I suggest, such diagnostic features are provided by the capacities of (1) cognitive control and (2) mental simulation subject to the requirement that both manifest explicit structure and flexible application. We'll start with the capacities.

The first, which corresponds to the "decentring" and "controlled" nature of the process, is executive. Whatever else an agent might do, if it's going to reason over something, it has to have some way of isolating relevant environmental features from distractors and maintaining relevant task parameters—including goals, rules, and strategies—over the course of the exercise. One can't deliberate about a chess move, for instance, if one can't attend to the relevant pieces or keep track of what moves are available or whose turn it is. To put the matter more technically, our capacity to deliberate draws upon mechanisms of executive attention to select relevant information and working memory to maintain it. Behaviorally, measures of both constructs correlate strongly with performance on problem-solving tasks (Engle, 2018; Simms et al., 2018), and neurobiologically, damage to their physiological substrates results in pronounced deficits in planning and long-term decision-making (Manes et al., 2002). These mechanisms and their contributions will be discussed in greater detail in section 3, but for now, suffice it to say that the two are standard in contemporary models of human reasoning (e.g., Koechlin, 2016; Eliasmith, 2013) and that, if other animals are to be deemed reasoners, they too will need some measure of control.

The second central feature, which covers the “entertaining alternative possibilities” aspect, is imaginative. Being able to isolate and maintain task-relevant information allows one to frame a problem, but coming up with an answer ultimately requires some means of constructing and evaluating potential solutions—of synthesizing the information one has. Depending on the problem faced, this can take a variety of forms. For some of the simpler problems studied in the lab, solutions may be found simply by combining the presence of an environmental feature (a pattern on a screen, say) with a task rule or heuristic (“if pattern X, do Y”; a concrete instance will be discussed in section 4.3). For more open-ended problems, however, it’s common to engage in a kind of “mental simulation,” playing through concrete sequences of events as a means of planning actions or figuring out how some mysterious event or process unfolded (Kahneman and Tversky, 1981; Klein, 1998; Johnson-Laird, 2010). Returning to the chess example, a player interested in maintaining a well-guarded king can envision various piece exchanges to see which move best accomplishes the goal. This strategy may not be the only one that one might wish to call reasoning, but it seems closest to the notion denied to animals by skeptics like Grotius and, as such, will be the main target here.

Now, supposing we take controlled imaginative inference as sufficient for reasoning, some qualifications are still in order. Many who doubt that animals reason in a human-like way are perfectly willing to grant that they engage in simple forms of inference, including forms that might satisfy the basic conditions above (e.g., Penn et al., 2008; Premack, 2010). If we are to vindicate the idea that animals have more in them than commonly imagined, we’ll need additional constraints. To this end, I have chosen two qualifiers invariably associated with “higher” cognition in the skeptical literature on animal reasoning. The first is “schematic,” a term used to describe how task variables are brought together. A schema is an explicit representational structure or template in which cognitive variables are embedded and made to play certain structurally defined roles (Rumelhart and Ortnoy, 1977; Gilboa and Marlatte, 2017).[45] The classic examples come from mathematics. The series A > B > C, for instance, offers a simple three-place structure in which variables (times, objects, people, etc.) are assigned roles as first, second, or third in the series. In the wild, schemas can, of course, be far more complex. A schema used to interpret the play of a game, for instance, might have roles for a dozen players corresponding to the myriad rules—formal and informal—that govern their conduct. Over the years, the idea has been modelled many different ways, taking on sentential (Cheng and Holyoak, 1985), imagistic (Johnson-Laird, 2006), and graphical formats (Gilboa and Marlatte, 2017). In each case, however, the authors have striven to capture the same core functional properties outlined above, viz. explicit structures transferrable across sensory particulars, the capacity to work with abstract variables, and role-differentiation.[46] That humans reason with schemas is widely granted, but many contend that other animals do not (Halford et al., 2010; Penn et al., 2008). They might use one stimulus dimension to predict another, associating brightness and reward amount, say, or generalizing between stimuli with the same perceptual features, but their inferential capacities would inevitably be limited to pairings between perceptibles learned from experience rather than abstract role-based similarity (see, e.g., Penn et

---

[45] The term "explicit" can mean many things. Concrete examples will come later, but broadly speaking, the present usage indicates merely that a relation or object maps on to some specific feature of the representational system. The greater than relation, for instance, may be explicitly represented by a ">" in a sentence or by the coordinates of a tensor.

[46] An independent line of investigation identifying the neural basis of schematic cognition in humans has recently been undertaken by Franklund and Greene (2020). Though the principal data on which they draw come from imaging studies, they are quite consistent with the picture of animal cognition developed here on the basis of recording data.

al., 2008). Concepts like "pawn," for instance, would have no significance, as pawns are defined by their in-game function rather than their looks (which may differ from set to set).

The second qualifier is "flexible" or, to put the matter more precisely, domain general. What is striking about humans, many argue, isn't just that we reason about complex concepts and relations but that—within certain limits (Carruthers, 2006, ch.6)—we do so across such a wide variety of contents. The same "tree" structures used to represent familial descent are used to organize evolutionary clades, business hierarchies, and traffic flows; the same capacity for imaginative simulation used to solve spatial problems is used also in the temporal and social domains—often all at once (e.g., in scheduling a conference). That other species do the same, even in a highly qualified manner, has long been subject to doubt (see, e.g., Premack, 2007; 2010). Evolution, it's observed, often gives rise to intricately structured but very narrow, domain-specific abilities—things like nest building in termites and wasps or, arguably, birdsong. There is little doubt that these behaviors are incredibly complex, skeptics note, but the complexity need not carry over into other aspects of life. Even if birds make use of a syntax in producing and parsing their song, they may not do so in thinking about time or space; even if macaques understand notions like "alpha" or recognize social hierarchies, this may be the only case where such capacities emerge. Animal cognition is "rigid," "specialized," or "single-purpose." If animals are going to be said to reason in a robust sense, the mechanisms will need to transcend such boundaries. At the very least, the mechanisms behind this schematic controlled imagination that we've been calling reasoning will need to work in *prima facie* different cognitive domains: social in addition to spatial, temporal as well as numerical.

4.2.2 Making a Neurobiological Case

The question now is whether other animals do satisfy these constraints and how we can know. Traditionally, arguments in favor of animal reason have been grounded on behavior. Tasks believed to require abstract or imaginative inference have been given, and animals' success taken as evidence. Yet as I argued at length in the previous chapter, such evidence has often been undercut. Critically, behaviors indicative of abstraction and imagination can result from processes that involve neither. Clearer indications, I have suggested, may be found by coupling these tasks and behaviors with additional, neurobiological measures. The present case offers a test of that claim. If I am right, present-day neurobiology makes a strong case for human-like reasoning in non-human animals. This isn't a terribly novel idea, of course. As we've seen, neural homologies have been a cornerstone of the animal reasoning case since the time of Huxley, Darwin, and Lewes. In its principal outlines, the method and argument made here will be one these figures would have recognized. When it comes to the neural substrates of "higher" reasoning, data from non-human animals form a pattern strikingly close to what is seen in humans and suggested by established psychological theory, and where parts and functions are similar, the mind should follow. Insofar as we accept some form of psycho-physiological parallelism, then, we should accept that animals reason much like ourselves. The principal novelty lies in the evidence. Whereas earlier arguments drew upon parallels in gross structure and function, tools made available in recent decades have produced a far more detailed picture of both human and animal neural processing. The development of imaging methods in humans and of recording and fine-grained interventions in animals have allowed a far better assessment of their relation than could have been imagined a century or even a generation ago. As the gaps in our knowledge have closed, I claim, it has become harder and harder to argue that human and animal minds fundamentally differ.

My argument concerns two crucial and interconnected populations: prefrontal cortex and hippocampus. These aren't the only regions involved, but they are among the best-documented, and given space constraints, provide as good a venue as any to address the issue. Years of neuropsychological, imaging, and, on occasion, recording studies have found that both regions play key roles in human reasoning (Kumaran et al., 2017; Stokes et al., 2017; Rubin, 2014). Broadly speaking, they can be thought of as complementary. In PFC and its subdivisions, we have, to a first approximation, the executive whose importance in task setting was emphasized in earlier. Damage to the area causes marked deficits in planning and rule-following, with lesion patients becoming easily distractible and facing well-known difficulties in guiding thought and action (Szczepanski and Knight, 2014). In the hippocampus and mid-temporal lobe, we have a set of mechanisms for more complex forms of inferential processing. Lesions there cause impairments to imagination (Hassabis et al., 2007), perspective taking (Beadle et al., 2013), and performance on transitive and other structure-based inferences (Moses and Ryan, 2006; Ryan et al., 2016). Imaging and intracranial recording evidence have lent additional support, tying regional activity to the navigation of social hierarchies (Tavares et al., 2015), the integration of events into narratives (Milivojevic et al., 2016), and certain forms of contextual linguistic processing (Piai et al., 2016). Activity in both regions correlates with working-memory intensive thought (Ranganath et al., 2004; Rissman et al., 2007), and as recent recording evidence from surgical patients shows, the two are tightly correlated during cognitively demanding tasks (Kaminski et al., 2017).

Unsurprisingly given the range and character of their functions, the regions have been taken as exemplars of schematic and domain-general processors (Gilboa and Marlatte, 2017). If fundamental, qualitative differences were present in human and animal reasoning, we would expect them to show up in how these regions work. Whatever it is about them that allows us to engage in deliberative and inferential thought should be absent or underdeveloped when we look beyond ourselves. Yet as I argue in the following sections, we find just the opposite. Human brains are larger and contain more recognizable sub-divisions than those of dogs, rats, or monkeys, but in terms of basic structure and operation, the regions appear much the same. Record from PFC or hippocampus while a rat solves a puzzle, and one finds precisely the structures and patterns one would expect and, to the extent that such information is available, find in human subjects.

4.3 Executive Capacities and the Prefrontal Cortex

We may begin by considering the executive functions embodied in the PFC. Anatomically, the region has long been taken as the locus of human uniqueness (Bianchi, 1922; Penn et al., 2008), in part because of its imposing size in our species. There remains little doubt, however, that the area has clear counterparts in others. That homologues[47] exist in other primates (Sallet et al., 2013), as well as animals like pigs, cats, and dogs is relatively uncontroversial (Fuster, 2015, ch. 2, ch. 4; Jelsing et al., 2006). Though it has been a point of controversy, rats seem to have it as well (at least according to some measures; Seamans et al., 2008). Though there is variation from animal to animal, regions identified as homologues tend to manifest specific structural features that, in the human literature, are cited as the foundations of higher cognition (Hunt and Hayden, 2017). It

---

[47] In the context of biology, a homology is a feature found across species and traceable to a common ancestor. In neuroscience, these are identified by several factors, including connectivity to other regions, cytoarchitecture, gross anatomy (location, relation to certain fissures, etc.), and functional profile as established by lesion and recording studies (see, e.g., Fuster, 2015, ch. 2).

houses dense recurrent connections, for instance (Goldman-Rakic, 1995; Wang et al., 2008), a feature thought to underwrite working memory retention. Likewise, it appears to be organized along a rough hierarchical gradient (Badre and D'Esposito, 2009; Alcaraz et al., 2016; Jin et al., 2016), a feature believed to allow the construction of schemas involving multiple levels of relation (e.g., analogies like "a is to b as c is to d" or "earlier is to later as near is to far"; Knowlton et al., 2012). The lesion literature seems to point in the same direction, as damage has been known to undermine animals' ability to stay on task and deal with multi-step problems. Lesioned mice show deficits in transitive inference (DeVito et al., 2010), while monkeys and dogs subject to frontal ablation are easily confused and may get lost during day-to-day activities (Bianchi, 1922). These features alone lend a bit of support to the idea that humans and animals with PFC homologues may think alike, at least at the most basic level—a comparable pattern of deficits on comparable tasks stemming from damage to comparably structured regions suggests a comparable mental process. We need not rely on anything so indirect, though. Fine-grained functional studies of purported rat and macaque PFC homologues accord remarkably well with the flexible schemas presented in section 4.2. Indeed, they resemble nothing so much as computational models of human executive function cited by skeptics themselves.

4.3.1 Computational Underpinnings

When thinking about how a neural system might give rise to the structured behaviors and thought patterns discussed above, authors have often appealed to a set of connectionist models labelled "integrated" or "compositional." Such architectures aim to ground the explicit schema and rule-based processing associated with classical symbolic systems in the neurobiologically inspired, tensor-based framework of connectionism (see Choo and Eliasmith, 2010; Halford et al., 2010;

Knowlton et al., 2012). Although they are not the only models available, they are perhaps the most successful in simulating complex relational and working-memory intensive thought. For example, the architecture developed by Eliasmith (2013) and his colleagues has been used to model the rule-based thought involved in human measures like the Wason Card Selection Task (Eliasmith, 2005) and the Ravens Progressive Matrices test (Eliasmith et al., 2012). Just as importantly for present purposes, they have been popular with those doubtful of animal rationality (e.g., Halford et al., 2010; Penn et al., 2008), who use the models to draw a qualitative distinction between the human and non-human mind. At some point in hominid evolution, it's argued, our brains diverged in such a way that they became capable of constructing schemas with role-based variables and explicit, generalizable structures. For the purposes of argument, I shall assume that human reasoning does involve something akin to what the compositional connectionist architectures present. My argument will therefore focus on the claim that this architecture type is uniquely human.

First, though, we need a bit more information regarding what these models entail. Unsurprisingly, they vary in their specifics. Some present themselves as "implementations" of classical symbolic approaches and others see themselves as alternatives, some make use of localist representations and others distributed, etc. These differences matter for how the models operate biologically and which high-level characterizations they incline toward (e.g., language of thought vs. mental model framings; Penn et al., 2008; Thagard, 2010). For our purposes, however, what's most relevant are their similarities. In each case, the architectures manifest a characteristic set of commitments that includes explicit components corresponding to the structures and roles defining a task. In approaches building on vector symbolic architectures, for example, complex representations like loves(John, Mary) are constructed using successive layers of neurons, each taking the tensor product or circular convolution of upstream assemblies. "Role" and "filler" ensembles respectively encoding the functional character (i.e., lover, loved) and sensory specifics (i.e., John, Mary) project to a collection of cells encoding their distinctive combination (e.g., John-lover, Mary-loved), these are, in turn, combined with ensembles corresponding to a relation (i.e., loves(lover, loved)) to yield the still more complex representation loves(John, Mary) at the next level (see, e.g., Eliasmith, 2005; Smolensky and Legendre, 2006). Thinking in terms of electrophysiology, models of this type would imply the existence of cells or ensembles (I remain agnostic as to which) that (i) signal the appropriate task relations, (ii) encode variables in accordance with the roles they play in a task, and (iii) combine these structures and variables into a single action-guiding signal. One can think of the process as one of problem-framing: task items are joined with relatively abstract schemas to suggest options.

The class of architectures attributed to animals by many skeptics are, by contrast, ones in which task items are assigned different values based on training history or where the agent learns to assess them along some stimulus dimension (Chen et al., 2014; von Fersen et al., 1991; Vasconcelos, 2008). Instead of employing an "explicit," recurring structure like "rule cells" or a serially ordered "list," the architecture encodes relations "implicitly" by directly modulating signals encoding unitary stimulus features. In Chen et al.'s (2014) model of magnitude judgment, for instance, individual animals are represented by a vector of characteristics like having a tail, living in water, and eating plants. These entries are, in turn, fed through a matrix of learned predictive weights connecting them to some other property of interest (e.g., largeness). This yields an estimated value for that property, and using basic signal detection theory, the estimates generated by different input vectors (animals) can be discriminated. One can, in other words, decide which of a pair of animals better exemplifies the property of largeness (defined as distance from an exemplar animal). In this way, the authors note, the system can make comparative *choices* without constructing an explicit comparative *relation* or schema like larger-than(a, b) to which individual animals are bound.[48] Thinking in terms of neural implementation, what one would expect if an account like this were correct would be cells that encode specific features (rather than generalizable roles) and are boosted or suppressed according to environmental associations between the property and the features manifested by the stimuli (instead of being "bound" to some separately encoded relation).

4.3.2 Evidence

With this bit of background in place, we may ask which kind of system the PFC in animals like macaques and rats more resemble. Recording investigations, I argue, reveal something much closer to the first. Over the past few decades, a growing body of evidence has amassed in favor of the region encoding roles, structures, and their systematic combination (see, e.g., Blackman et al., 2016; Cromer et al., 2010; Tsutsui et al., 2016; Wallis et al., 2001). At the same time, considerable doubt has grown as to the plausibility of stimulus feature encoding (Duverne and Koechlin, 2017; Lara and Wallis, 2014; Stokes, Buschmann, and Miller, 2017). The more we find out about the area's operation in humans and other animals, it seems, the further we are pushed from the exceptionalism endorsed by skeptics.

We can begin with roles. The idea that the PFC may employ functional rather than featural categories has been around for some time, gaining indirect support from the sheer range of stimuli with which the region operates. The PFC has been implicated in tasks involving odors and textures

---

[48] What *is* explicitly represented in this model are the animals (by their feature vectors), the properties evaluated (e.g., largeness), and the predictive relations mapping the former to the latter.

(Bissonette and Roesch, 2015), colors and orientations (Buschmann et al., 2012), numbers and magnitudes (Nieder, 2016), and many more categories. Yet, during a given task, as many as thirty percent of recorded neurons will carry information about a specific task variable (Duncan and Miller, 2002). Were these cells encoding specific perceptible features instead of abstract roles, the argument goes, the region simply wouldn't have enough space to cover the range of items it does. Accordingly, many have assumed that the region must be involved in a more flexible, multi-task form of processing (Stokes et al., 2017). The last ten years or so has seen a large body of evidence to support this, with some of the most critical findings coming from studies involving the transfer of roles between different stimulus types. Examined in the context of a single experimental setup, where an individual or feature-based category (e.g., John) always plays the same role (e.g., loved), the responses of feature and role-encoding cells will covary, making the competing architectures difficult or impossible to distinguish. Examined across different settings, though, the two can come apart. The same task can be run with stimuli instantiating very different perceptual profiles, keeping the roles fixed and varying the features; conversely, a different task might be run using the same stimuli. To obtain evidence for role-based over feature-based encoding, then, one must find cases where cells or ensembles track roles over percepts in such dissociating cases. The opposite should hold for a feature-based theory.

Cases involving manipulations like those described above have been run with both rats (Rich and Shapiro, 2009) and macaques (Cromer et al., 2010), and in each case have revealed a pattern of evidence favoring role over feature-based encoding. Cromer et al. (2010), for example, tracked a population of lateral PFC cells during two-group categorization tasks run with images of cats vs. dogs in one condition and sportscars vs. sedans in another. Consistent with a role-encoding architecture, they found that exact same cells tended to distinguish very different sets of stimuli:

if a cell had a clear preference in the animal condition, increased firing for cats, say, then it most likely had a similar categorical selectivity in the cars case. Assuming the perceptual features differentiating cats and dogs differ from those dissociating sportscars from sedans, however, the cells would appear to be tracking the objects' generic functional significance—that they belong to different categories—rather than their "cathood" or "sedanhood."[49] Still more strikingly, Andreas Nieder and his colleagues have reported cells in the macaque PFC and parietal cortex that seem to signal numerosity (see Nieder [2016] for a review). These "number neurons" display clear preferences for the number of items with which the animal is presented, with cells covering a documented range spanning from zero (Ramirez-Cardinas et al., 2016) to as high as 30 (Nieder and Merten, 2007). As in the case of categorical judgments, the critical findings concern the wide array of perceptual stimuli to which the cells are responsive. Macaques have had to match visual patterns across variation in the size, color, shape, and spatial arrangement of the objects pictured (Moskaleva and Nieder, 2014), pair simultaneously presented numbers of dots with sequential flashes of light or bursts of sound (Nieder, 2012), and even generalize operations learned with dot patterns to arbitrary symbols like "4" (Diester and Nieder, 2007). In each case, the activity of the neurons was stable across conditions. Were these cells tracking a perceptual feature like stimulus entropy rather than an abstract role-based concept, one wouldn't expect so broad and consistent a range of activity. Indeed, one might not expect distinct cells to signal numbers at all. Were we to adapt Chen et al.'s (2014) model to the task, for instance, we would expect dot pattern comparisons

---

[49] Curiously, Knowlton et al. (2012) cite this study as a support for the biological plausibility of their model. It's unclear how the lab means to square the use of this evidence with their claims elsewhere that non-human animals lack such cognitive architectures.

to be made using cells that monotonically scale their activity up (or down) based on the density of the pattern instead of having an array of neurons with regularly occurring peaks.

A similar story can be told for the relations that hold between such role-filling variables. As I mentioned earlier, skeptical authors are perfectly willing to grant that animals make relational discriminations, provided the mechanism is feature-dependent and encodes structure “implicitly” in the associations or values attached to these features. The ability to press a button whenever two objects are the same size, for instance, could easily be mastered by the architecture designed by Chen and colleagues. What’s denied are architectural features “based on the structural relation between abstract functional roles” (Penn et al., 2008, p. 114). The architecture should not have anything so “explicit” as the role-fillers and rules posited by compositional connectionist models. Experimentally, one would expect the difference between the explicit and implicit architectures to show up in a few ways. The most obvious is that the former will have components that, when examined across shifts in task rules and stimuli, tend to track specific relations. When judging whether two items are the same, for example, one would expect the explicit architecture to contain elements that signal “sameness” over “difference” and vice versa, are selectively engaged during similarity judgment tasks, and can be “bound” to features and stimuli from different perceptual domains. For the implicit models, comparisons are accomplished by something less involved—applying a general signal detection mechanism to activity stemming from feature-sensitive cells, for instance (Chen et al., 2014). A population might signal when the stimulus variability of two cues cannot be distinguished (something like this is suggested by Penn et al., 2008, p. 112), but one wouldn’t expect distinct “sameness” and “difference” signaling populations that carry this signaling pattern between features (color, shape, etc.).

Here again, the electrophysiological literature seems closer to the explicit relational structures. Sensory-feature independent “rule cells” have been reported for nearly two decades (Wallis et al., 2001), including in humans (Mian et al., 2014), macaques (Buschmann et al., 2012), and rats (Bisonette and Roesch, 2015; for review, see Bissonette and Roesch, 2017). As their name suggests, the cells characteristically signal when a specific set of relations are relevant for a task. Examples include “greater-than” and “less-than” (Bongard and Nieder, 2010; Eiselt and Nieder, 2013), “sameness” and “difference” (Hussar and Pasternak, 2013), and various “if-then” contingencies (Tsutsui et al. 2016; Wallis et al., 2001). In each case, the cells seem to follow the relation over anything else. If the same rule is signaled by different cues or brought to bear on different role-fillers, activity in “rule”-encoding populations remains fixed (Eiselt and Nieder, 2013; Hussar and Pasternak, 2013; Shima et al., 2007; Wallis et al., 2001). For instance, Hussar and Pasternak (2013) report that 74% of cells signaling “sameness” and “difference” during speed discrimination played the same role when their subjects were comparing directions. Similarly, Eiselt and Nieder (2013) find that 34% of cells signaling “greater-than” and “less-than” during number comparisons do the same when assessing magnitudes, a variable distinct from numerosity and represented by distinct cells (see Moskaleva and Nieder, 2014). When one examines tasks that require switching between multiple rules instead of objects, the opposite tendency emerges. Rules are suppressed or facilitated depending on what is currently relevant for the task (Buschmann et al., 2012; Eiselt and Nieder, 2013; Powell and Redish, 2016; Tsutsui et al., 2016).

Consistent with the explicit rule- and role-based models, the two combine in systematic ways to guide behavior. In the case of motor sequence planning, for instance, rule cells corresponding to the execution sequence ABAB and AABB only lead to specific actions when combined with cues indicating which movements fill the A and B slots: “push” and “pull” might be combined with ABAB to signal to “push pull push pull” but the same rule could also give rise to “turn push turn push” given different cues (Shima et al., 2007). Depending on the role-fillers and their positioning, the same rule can thus lead to different behaviors. The converse is also true. Given one has learned a set of functional categories, one should be able to pair them with many different rules. The same array of number neurons, for instance, may be paired with greater-than, less-than, same-as, different-from, and potentially many more operations (e.g., addition, a behavioral capacity observed in both macaques [Livingstone et al., 2014] and passerine birds [Pepperberg and Carey, 2012]). The specifics of how role-fillers and relations are bound together remains a topic of investigation, with evidence pointing toward at least two potential mechanisms.[50] However it’s accomplished, though, the outcome appears to be the same. Signals playing the functional roles of variables and relations are bound together in such a way that they produce simple inferences relating to potential actions or strategies.

This section has given us a lot to keep track of, but a recent study by Tsutsui et al. (2016) provides a nice illustration of how the parts—the abstract roles, explicit rules, and their combination—work in practice. Macaques were trained to anticipate fruit juice delivery based on one set of random images and saline based on another. Midway through, the experimenters reversed the contingencies, tricking the macaques into an unexpected mouthful of saltwater. Immediately following the surprise, the experimenters saw a change in the subjects’ behaviors and drastic shifts in rule cells and outcome-predictive neurons but not the category-sensitive ones. Prior

---

[50] Specifically, there is evidence for both “binding by synchrony,” in which rules and role-fillers are brought together by synchronic activation in their respective populations (Buschmann et al., 2012), and “binding by convergence,” where the two feed into a third population of neurons representing their conjunction (Tsutsui et al., 2016; Del Arco et al., 2017).

to the reversal, “A” cells paired with cells signaling the rule set $(A \rightarrow X) \cdot (B \rightarrow Y)$ gave rise to activity in downstream “response contingency” neurons signaling fruit juice (i.e., X); afterward, the same cue combined with the new set $(A \rightarrow Y) \cdot (B \rightarrow X)$ to trigger saline-predictive activity (i.e., Y; see Del Arco et al. [2017] for similar results in rats). The category sensitive neurons, the authors surmised, were functional rather than featural, since the use of random images and the rule reversal eliminated any perceptually based properties they may have shared. The rules these categories were paired with were both explicit, represented by distinct and separable cells, and flexible, encompassing categories they had not been trained on. Finally, the way in which the two were paired appeared to shape the responses of downstream action-guiding populations in a systematic way. Indeed, when the three cell types’ signals are listed in order of firing, one gets a canonical modus ponens: $(A \rightarrow X); A; \therefore X.$ [51] Doctor Preston may well have been right to credit the dog with a syllogism, then.

4.3.3 Objections

Taken at face value, these electrophysiological results suggest that not only humans but monkeys, rodents, and other animals with developed PFCs make use of flexible, abstract, and rule-based cognitive architectures. The categories may not be as abstract as those used by humans and the rules may not be as deep; in terms of basic problem framing and inferences, though, they seem

---

[51] This linguistic notation, it should be emphasized, has been used only for convenience. The core of the propositions could, so far as I am aware, be rendered just as easily in graphical or imagistic formats (see Hughes, 2006; Johnson-laird, 2010). While these results suggest that non-human thought is systematic, then, they need not imply anything like language of thought.

to be going about tasks in a "human-like" way. But appearances can be deceiving. Before turning to consider the imaginative component of reason, then, we need to consider what the alternatives might be. I am not aware of any discussions of the above-cited results by skeptically inclined authors, but a good sense of how one might respond can be obtained by observing some longstanding objections from discussions in the behavioral literature. Traditionally, when behavioral evidence has been presented in favor of abstract, rule-based executive processing in animals, one of two problems have been raised. The first is that there could be confounds in the stimuli used. An animal might not be responding to the instantiation of an *abstract* concept but to some uncontrolled stimulus dimension. The second is that, even if they exist, the rules and roles that animals understand may not be sufficiently general or flexible in use. Most studies, including the majority of the neurobiological investigations considered here, do not test for generalization. Given the controversial nature of the claims, then, one might desire more substantive evidence of rule transfer than has been common.

To the first concern, I would offer three lines of reply, two empirical and one conceptual. First, there is a large measure of implausibility to the assertion that these animals are tracking some stimulus dimension when one considers the range of stimuli to which these cells respond. Take the case of number neurons. To effectively deflate the interpretation given by Nieder and his colleagues, one would need to show that there was some perceptual feature that accounted at least as well for the cells' responses as abstract numerosity. Were the studies limited to dot patterns, this may not be difficult, but the fact is that the same pattern of cellular response has been observed in the case of tones, flashes, dots, and symbols. There may be a common feature here (see below), but none has been presented. Second, there is the observation that signals recorded in PFC often differ substantially from those recorded in upstream perceptual regions (Cromer et al., 2010;

Ramirez-Cardinas et al., 2016). Were the representations understood by animals sensory in nature, as our interlocutor suggests, one would expect them to conform to patterns set by other known perceptual signals, such as those obtained from populations tracking textures or colors. They don’t, though. Once again, the most obvious examples are prefrontal numerical representations, which are unique in, among other things, their insensitivity to modality, their responsiveness to arbitrary symbols, and their number-like encoding of blank stimuli (i.e., zero; Ramirez-Cardinas et al., 2016). In short, they don’t look like perceptual representations. Outside prior theoretical commitments, then, there’s no clear reason to lump them together.

But suppose one were to find some sensory re-description of the information carried by these cells. Even then, I would argue, we need not assume that they cannot count as abstract categories. The philosophical literature is replete with properties like “grue” (Goodman, 1983) and rules like “quus” (Kripke, 1982) that cannot be ruled out by existing data; it likewise contains many discussions of coextensive proximal and distal properties and the problems they pose when ascribing mental contents (see, e.g., Millikan, 2004). The specifics cannot be discussed here, but both literatures show at least two things. First, re-descriptions are not a problem limited to animal cognition. The exact same objections raised by our skeptic could be raised with regard to evidence drawn from human subjects, rendering their selective use in the case of animals a clear double standard. Second, the mere existence of alternative interpretations does not suffice to decide the issue or, in cases like green and grue, undermine commonsense readings. At the very least, if the skeptic means to adopt a sensory-feature reading instead of the common reading given in comparable human cases, they will need to offer a substantive defense (which may prove difficult, see Buckner [2014]).

Regarding generalization, I would note two things. First, we should not undervalue the evidence we have of generalization. I briefly mentioned two studies, those of Eiselt and Nieder (2013) on magnitudes and numbers and Hussar and Pasternak (2013) on speed and direction, that purport to show generalization between domains. The reason this is particularly impressive—and a point particularly worth emphasizing at the present moment—is that in both cases the stimulus features relevant to the assessment of the two properties can and did come into conflict. A fixed number of items can vary in size and items of the same size can vary in number; objects can move in the same direction at different speeds and vice versa. To make an accurate greater-than or less-than assessment in the case of shrinking and growing dots requires one to ignore their diameters, while doing the same for the magnitudes of two bars requires ignoring the fact that there is only one bar in both stimuli (*mutatis mutandis* for movements of varying speeds and directions). Were the relations evaluated in these cases perceptually embedded, we would expect separate size and magnitude specific greater-than/less-than and distinct speed and direction matching mechanisms. The fact that sizeable numbers of rule cells operated across contexts suggests a more general and flexible notion: at the cellular level, there appears to be a higher-order commonality or analogy drawn between the various ways two things can be the same, different, greater-than, and less-than.

Second, the hypothesis that the PFC flexibly re-deploys relations in the way suggested by these studies helps to explain many other findings. For one, there is a large body of complementary evidence for generalization from the behavioral literature. These are easily accounted for by positing a flexible mechanism but, as simulations show, are not predicted by standard deflationary architectures (Jensen et al., 2013; Mininni and Zanutto, 2017). On top of this, we have good abductive reason for thinking the observed relations generalize. There are quite a few domains that seem to be organized very similarly in terms of their operation in PFC and that are subject to the same or analogous rules (ranking, similarity, etc.). For example, the functional organization underlying macaque numerical cognition—a "number line" of lateral prefrontal cells with peak responsiveness for individual numbers and overlap with nearby values—is mirrored in the organization of magnitudes (Eiselt and Nieder, 2013), ratios (Vallentin and Nieder, 2010), temporal ordering (Tiganj et al., 2018), spatial frequencies (Eiselt and Nieder, 2016), and transitively ranked reward values (Brunamonti et al., 2016). Each is a domain in which comparisons of things like ranking naturally occur and in which cross-domain comparisons would likely prove fruitful (e.g., in assessing the tradeoff of time vs. reward). For flexible schematic models, the existence of these similarly organized domains makes perfect sense. Each reflects one application of a more general schema and rule set. For deflationary accounts to make sense of the parallels, though, such commonalities would have to be either accidental or the result of a common perceptual feature or training history. The first option would end up positing six co-localized and curiously similar organizational structures when one would do. The second would have to find something similar in each case, whether a sensory feature or a hidden reinforcer. Yet the studies from which these observations have been drawn involved *prima facie* very different domains, standardly controlled for training history (see, e.g., Brunamonti et al., 2016), and in some cases never reinforced the relevant task dimension at all (Tiganj et al., 2018).

4.4 Imagination and the Hippocampus

Supposing that non-human animals do isolate task variables and rules, the question becomes how they are used. We have evidence of PFC-mediated inference in simple cases of action selection, but it's doubtful these mechanisms scale up. As more choices become available and as the proper strategy becomes more and more ambiguous, rule and variable combinations will rapidly outstrip working memory and the prefrontal mechanisms that ground it. Even under ideal conditions, no mortal will ever deduce the optimal moves in chess simply by following the rules and piece arrangement to their logical ends. Something else is needed. As descriptive accounts of human reasoning have noted for some time, the norm in messy, real-world reasoning isn't a straightforward process of deduction but a kind of targeted mental exploration (Johnson-Laird, 2010; Simon, 1956; Kahneman and Tversky, 1981). Rather than solve a task analytically, one simulates or imagines. Such simulations may take place across cortical regions (Hesslow, 2012), but the hippocampus and surrounding temporal cortex appear especially important (Rubin, 2014). As was mentioned earlier, individuals who've suffered hippocampal damage manifest difficulties in activities requiring precisely these capacities. In terms of computations, these imaginative inferences are believed to build on the cognitive maps discussed in chapter 3 (Eichenbaum, 2017a; Milivojevic et al., 2016). By bringing the construction and operation of these rich representations under direct cognitive control and by developing the ability to reuse relational structures in spatial to non-spatial domains, humans are believed to have developed the capacity for flexible, offline mental simulation (Suddendorf et al., 2014).

The problem for opponents of animal reasoning, however, is we find the same structures and processes in non-humans. Starting with basics, we know that hippocampus is a highly conserved structure, evolutionarily speaking, and that it manifests the same architectural properties in other species that are believed to underwrite its inferential processing functions in humans (the same recurrent connections, hierarchical gradients, etc.; Eichenbaum, 2017a). We know also that damage to the region produces deficits analogous to those observed in humans. Lesioned rats encounter the exact same difficulties in drawing transitive and transverse (A implies B implies C implies A) inferences as humans who've suffered similar damage (Dusek and Eichenbaum, 1997; 1998; Cf. Ryan et al., 2016) and manifest impairments across a comparably wide array of cognitive domains, including in temporal, social, and other non-spatial tasks (Schiller et al., 2014). Once again, though, the strongest evidence comes from studies at a finer grain.

4.4.1 Principles of Prospective Imagination

Around this point in section 3, we paused to consider the computational underpinnings of the capacity served by the PFC, drawing a few implications from its diverging models and using them as a yardstick for the electrophysiological data. The same cannot be done in the case of imagination, though not for the reasons one might expect. The problem isn't that imagination is inherently murky or less amenable to computational modelling. Though less established than work on rule-guidance and relational cognition in the PFC, there exist plenty of computational accounts of how a neural system might give rise to simulation (e.g., Chalmers et al., 2016; Pezzulo et al., 2017; Momennejad, 2020). Were one so disposed, such proposals could be used as a guide in the same way as the explicit rule-based models were earlier (we'll see a few specifics in 4.4.2). The problem is that such comparisons would most likely be circular. Present-day neurobiological models of imagination draw explicit inspiration from experimental results in the animal literature and in many cases are constructed with the aim of matching evidence from rat studies. This situation bodes well for the case that other animals imagine but trivializes any match we might find between the electrophysiological results and such hypotheses.

What we can do instead is consider more general principles and properties. Specific computational ideas aside, what qualitative features would one expect from a psychological mechanism tasked with simulating actions and outcomes? A useful breakdown is provided by

Suddendorf and Corballis (2007). By way of characterizing the richer notion of Mental Time Travel, the authors list several components that appear necessary for the prior and ultimately simpler notion of mental simulation. Taking the theater as their metaphor, they note that a play cannot be staged without, among other things, a venue, a set, actors, and a playwright or narrative.[52] The venue in the case of imagination is working memory, which provides a space for a variety of productions, including dreams, day-dreams, mental rehearsals, and most importantly from the perspective of decision-making, a form of thought labelled "secondary representation" linking present conditions to distant states of affairs. The set, for its part, includes our knowledge of physical objects and their characteristic behaviors (e.g., the shapes and actions of various chess pieces) as well as a sense of time, or how these behaviors may be arranged. The actors are constituted by our knowledge of ourselves and others (characteristic behaviors, likely choices, etc.). Finally, the playwright draws all this disparate information from long-term memory and combining it in a such a way that it forms a coherent whole.

In terms of neurobiological underpinnings, it's not clear that these terms necessarily imply any one form of organization. There are several broad characteristics that one would expect, however. Minimally, one needs elements of the cognitive architecture, whether cells, ensembles, or networks, that encode the "contents" of the imagining. One would expect to find signals corresponding to different set pieces, for instance. One should also be able to pinpoint the actors,

---

[52] They also include a director and executive producer, who play roles associated with the PFC, and a broadcaster that communicates events to a wider audience. These functions may be bracketed for the present discussion, however, as the first two have previously been discussed and the last isn't strictly necessary for inference to take place (a person could be wholly paralyzed without losing their ability to consider what they would do if they weren't).

however many there are, and trace their actions over the course of the episode. Each of these elements should, in turn, be embedded within some overarching relational structure. Objects should stand in a determinable spatiotemporal position relative to each other and/or to the actor(s). Finally, the mechanism must construct and deploy these imagined scenarios in a way that ultimately aids in decision-making. Given a problem believed to prompt such thinking, one should be able to observe the system as it works through potential scenarios—going "this way or that" at a fork, say. Each of these characteristics, I argue, can be found in non-human hippocampal processing.

4.4.2 Evidence

I'll start with the basic materials. Without re-hashing all the material from the previous chapter, we may note a large body of evidence supports the idea that hippocampal activity in humans and other animals is organized into maps: schematic structures that bind together spatial, temporal, and other forms of information into complex networks (Epstein et al., 2017; Rubin et al., 2014; Schiller et al., 2014; for a philosophical discussion, see Bechtel, 2016). Task variables are mapped onto specific cells, while between-variable relations are captured by spatiotemporal firing patterns at the level of ensembles (Chen et al., 2015). If the domain is spatial, for example, one would expect individual cells to take on spatial firing fields that, taken together, reflect the structure of the environment the subject is tested in. If the domain is relatively stable, these structural relations and mappings can be maintained and matched over time and multiple exposures (Sellami et al., 2017). If the map ceases to be accurate, they can be updated, either all at once—rotating if the room rotates, say, or stretching if it stretches—or narrowly and in a piecemeal fashion—as when a new barrier or object is encountered (Derdikman and Moser, 2010). In humans, these properties are widely believed to underwrite the basic staging role highlighted by Suddendorf and Corballis. The ability to organize large amounts of information in a systematic manner, the capacity for selective updating and transformation, and the many findings linking the hippocampus to narrative construction and memory all make the structure a natural candidate (Chen et al., 2015; Cohn-Sheehy and Ranganath, 2017; Epstein et al., 2017; Rubin et al., 2014). In cellular recordings with human surgical patients, for instance, variable-specific single neuron activity analogous to "place cell" activity in rats has been observed during the recall of specific locations, the visualization of objects and people, and even certain forms of context-sensitive language processing (Miller et al., 2013; Piai et al., 2016; Jafarpour et al., 2017). Likewise, multivariate fMRI analyses have found evidence of systematic "mappings" in a variety of contexts, including the targeted recall of real-world times and locations (Nielson et al., 2015), the integration of multi-episode fictional narratives (Milivojevic et al. 2015; 2016), and the interpretation of sentences based syntactically organized semantic information (e.g., anticipating the end of the sentence "He sailed in a wooden…"; Jafarpour et al., 2017). Given a top-down signal from the PFC to cue a memory or scenario, the region appears capable of filling-in myriad unstated structural details and using them generate plausible predictions (the timing of an event, the end of a sentence, etc.).

The question is whether other animals use their maps in this way. The answer seems to be yes. Though first documented in the context of location-tracking, variable-specific cellular encoding has since been observed for a variety of "set pieces." In rodents, cells have been found to encode not only the animal's location but the location of goals (Gauthier and Tank, 2017), the presence of environmental barriers and direction of motion (e.g., whether a track is one-way or two-way; Moser, 2015), the time when an event or object was encountered or may be expected

(Eichenbaum and Cohen, 2014), and the sensory specifics and valence of individual objects (e.g., smell and reward history; McKenzie et al., 2014). A recent letter to *Nature* even reports the formation of an "auditory map" in rats trained to "locate" frequencies using a joystick (Aronov et al., 2017). In the more visually attuned macaque, there is also evidence that hippocampal cells encode subject point of view and can anticipate which landmarks and features will become visible from a given location (Wirth et al., 2017). In terms of "actors," cellular recordings from both mice and macaques indicate that the region encodes not only the position of the animal themselves, but the presence of individual conspecifics (Okuyama et al., 2016; Sliwa et al., 2016). Consistent with recordings in humans (Quiroga et al., 2005), these cells appear especially attuned to familiar conspecifics. In the case of mice, moreover, there is evidence that the inception of "false memories" by optogenetically coactivating such neurons with negative or positive stimuli (i.e., foot shocks or cocaine administration) causes selective shifts in approach and avoidance behaviors directed toward those individuals (suggesting the activity isn't epiphenomenal; Okuyama et al., 2016). Just as in humans, then, the hippocampus of monkeys and rodents appears capable of tracking a multiplicity of variables—including objects, locations, times, and actors—and binding certain properties like valence and visibility from a location. If brought together, these various items would prove more than enough to stage a simple play.

But can they be brought together in a systematic way? Again, the answer seems to be yes. Looking to the literature on mapping, one can find all manner of arrangements between spatial and non-spatial elements. There are hierarchies like forking paths and one-way linear series (e.g., a musical scale; Aronov et al., 2017), ring structures like circular tracks or the transverse series A>B>C>D>A (which humans, macaques, and rats with hippocampal damage struggle to learn; Alvarado et al., 2002; Dusek and Eichenbaum, 1998; Rickard et al., 2006), and countless other relational structures (figure eights, two-way tracks, etc.). What's more, available evidence indicates that these spatial and non-spatial elements regularly combine with one another to form multi-dimensional representations. The above-cited macaque study by Wirth and colleagues, for instance, found cells that simultaneously encoded the animal's position in space, direction of gaze, and the presence or expectation of landmarks, while McKenzie et al. (2014; 2016) report that populations in the rodent hippocampus encode not only the locations of objects, but their properties, valences, and task roles. What gets represented and how appears, in turn, to depend on the contingencies of the task at hand (something signaled by inputs from the PFC; see Eichenbaum, 2017b). One sees a hierarchy of task components depending on their importance. If finding food requires choosing a container based on the kind of room it's in, say, the hippocampal representation of different rooms will be further separated in coding space than the representation of different containers (McKenzie et al., 2014); if timing is critical, it will be the principal coding dimension (Allen et al., 2016; see Milivojevic et al. [2015] and [2016] for the human case). Cognitive maps aren't just a tool for plotting location, then, but a flexibly deployed panorama of times, places, objects, and their use to the viewer (Schiller et al., 2014; Lisman et al., 2017). In other words, non-human animals not only have the raw materials for the show but the venue and organizational capacity required to stage and narrate it.

This brings us to the final question. Are these structures ever actually used to simulate actions? Is there a narrative? Abstractly, they seem like the right tool for the job. Recurrent neural networks—of which the hippocampal CA3 and CA1 subregions offer the most famous biological example—are commonly used for executing simulations in AI (Chiappa et al., 2017) and, as a class of network, are known to be universal dynamical systems approximators (i.e., capable of simulating any system that predictably evolves over time; Bianchi et al., 2017). They also fall quite naturally from the best existing models of the cognitive map discussed in chapter 3. At present, it's standard to represent the network of "place" cells that compose the map as transition matrices, tensors encoding the probability of moving from one state to another given a set of actions and an increment of time.[53] Among other things, these nicely account for the response fields of cells in the hippocampus (Rikhye et al., 2019), their activity during sleep (Cazé et al., 2018), and their relation to the well-known "grid" cells found in upstream entorhinal cortex (the firing properties of which correspond to eigenvectors of the posited hippocampal transition matrix; Stachenfeld et al. 2017). Most importantly for the question of simulation, they also make playing out future contingencies relatively easy. Starting from state s_0, one can simulate trajectories by recursively feeding predicted states and potential actions through the transition matrix (Mattar and Daw, 2018; Stoianov et al., 2018; Rikhye et al., 2019)—a process readily implemented by the aforementioned recurrent connections within the hippocampus (supplemented by strategy-directing inputs from PFC).

The real evidence, however, comes from the recordings. In line with the picture sketched above, the hippocampus has been known for decades to house what are sometimes called internally generated sequences (IGSs), patterns of activity that break from the animal's present location and trace their way through distant regions in the cognitive map. IGSs fall into two broad categories: theta-frequency "vicarious trial and error" sequences (VTE), which are commonly observed during

---

[53] Formally, the matrix T will contain entries corresponding to the probability of going from one state to another in a single time step, conditional on some action. A particularly striking image of population activity for a given state (location) may be obtained by plotting the corresponding column of matrix M, the discounted n-step sum of T under a given action policy: $M = \sum_{n=0}^{\infty} \gamma^n T^n$ (i.e., the "successor representation"; Stachenfeld et al., 2017).

extended periods of decision-making, and sharp wave ripple precessions (SWRs), a faster-paced sequence associated with both post-exploration rest and in-the-moment decision-making. The phenomena are still under investigation, but the standard run of events goes something like this. To start, increased communication is initiated between PFC and hippocampus (Redish, 2016), the former playing an initiating or question-framing role (Schmidt et al., 2019). In the context of maze-running, for example, the PFC might signal a distant goal position or the animal's distance to the goal, switching as the task contingencies change (e.g., from left-right to cardinal direction-based goal assignment; Kaefer et al., 2020). If the rat is currently active, it will pause and the hippocampus will switch from extrinsic sensory drivers to intrinsic "offline" processing (if the rat is resting post-exploration, it will be offline to start with; Colgin, 2016). Like someone trying to figure out whether to take the freeway or go through the city, the rat then works through its options. Place cell by place cell, sequences trace paths down the corridors of the cognitive map, going in one direction or another depending on the environment and the task they're performing. In the case of a simple T-Maze, activity will go down one arm and then another (Johnson and Redish, 2007). If the goal is further away, the sequence will go longer (Wikenheiser and Redish, 2015), and if the rat pauses after making a wrong turn, it will trace its way backward instead of going ahead (Johnson and Redish, 2007). As activity makes its way through these distant states, it naturally calls up information associated with each (e.g., positive or negative outcomes; McKenzie et al., 2014), relaying these predictions to "evaluative" regions like the ventral striatum, amygdala, and orbitofrontal cortex where they may be contrasted with the outcomes of other options (Redish, 2016; Wikenheiser and Schoenbaum, 2016).

The discovery of these sequences has led more than a few psychologists and neuroscientists—including once-skeptical authors (e.g. Corballis, 2013; 2017)—to characterize them as neural correlates of deliberation or episodic reasoning (Johnson and Crowe, 2009; Redish, 2016; van der Meer et al., 2012). It isn't hard to see why. What I've called imaginative inference is traditionally thought of as a "decoupled," "decentered," "serial," "controlled," and "goal-directed" process (Evans and Stanovich, 2013; Hurley and Nudds, 2006) or the activity of "envisaging the possibilities consistent with the starting point" (Johnson-Laird, 2010, p. 18243). In VTE and SWRs, such terms are given concrete meaning. Activity in the hippocampus literally *decouples* from immediate sensory inputs, locking onto prefrontal signals and decentering from the animal's present location. Faced with a choice between corridors, this activity *serially* traces routes from a starting point of the animal's current location down alternate paths, entering into distant and frequently imperceptible locations (e.g., areas behind corners). Where these paths go is determined not simply by the immediate stimuli the animal receives but by the task it is performing, with differing patterns of mental exploration emerging based on the animal's *goals* (Johnson and Redish, 2007; Singer et al., 2013; Kaefer et al., 2020). Computationally, the system seems to be performing a simulation in a very traditional sense of the word: sampling from a state transition model that mirrors contingencies in some domain of interest (Foster, 2017; Mattar and Daw, 2018; Stoianov et al., 2018; see Xu et al., 2020 for an abstract treatment). Intuitively speaking, it looks like it's imagining the likely outcomes of the choices available.

4.4.3 Objections

We have evidence, then, not only for a system of organized relational knowledge but a process of guided inference taking place within it. To my knowledge, this much has not been challenged. There are, however, some worries regarding how much can be inferred from these observations. One concern, expressed by Suddendorf (2013a), is that these inferences may be limited in scope. He emphasizes the richness of human imagination—the sheer number of actors and actions it encompasses and its remarkable ability to embed events in larger narratives (p. 151)—and suggests that the internally generated sequences observed in rats may amount to little more than immediate trajectory planning. Similarly, Cheng et al. (2016) have contended that these sequences might be understood as "memory traces" of specific experiences rather than flexible constructions. The ability to select actions based on recalled episodes would count as a kind of inference, they allow, but it would be a rather limited kind. Human imagination, they note, operates not just by encoding and recalling episodes but by "combining and recombining basic elements into novel constellations" (p. 85). They grant that some studies, such as Gupta et al.'s (2010) work establishing that rodent SWRs pass through novel combinations of regions, challenge this explanation. But they caution that the function and proper interpretation of the sequences remains unclear, citing simulation results indicating that similar unexperienced sequences can be generated randomly in a neural network. Though suggestive, they argue, such results are inconclusive when it comes to "higher" cognition. To use the terminology of the present chapter, the concerns are that the inferences may be inflexible or bound to a specific experience (and therefore non-schematic).

There are several facts and findings that speak against these objections. To start with, there is good evidence that these sequences are not limited to familiar episodes. Were it the case that VTE and SWRs represented a kind of "video replay" for events, as the objection implies, we would not expect them to extend to regions or areas that the animals have never been. This much is clear from how they address the Gupta et al. study. Yet work by Ólafsdóttir et al. (2015) and Mou and Ji (2016) among others make a strong case that they do, and in both studies, the counter that the results are hard to interpret or may emerge randomly proves much less plausible. The first involved exposing rats to novel maze in which a large portion was blocked by a transparent barrier. Rats were allowed to explore unblocked regions, but when subsequent SWRs were examined, the authors found that they extended well into zones inaccessible to them. What's more, they discovered that sequences could be heavily biased toward a specific location in the blocked zone by having rats watch an experimenter deposit food there after they'd left the maze. In other words, the sequences were goal sensitive. More impressively, Mou and Ji (2016) report that VTE-like theta sequences emerge when rats watch conspecifics run on a separate, inaccessible, and unfamiliar track, with the same sequences being observed in viewing as later arose when rats were placed on the track themselves. Intuitively speaking, the rats put themselves in the other rat's place—a process that facilitated their own learning and did not occur with several distinct controls.

In both cases, the sequences' functions are reasonably clear. In Mou and Ji's study, rats who engaged in VTE were faster learners, while studies interrupting SWRs—the sequences observed by Ólafsdóttir et al.—have consistently found deficits in subsequent decision-making (Girardeau et al., 2009; Pfeiffer and Foster, 2013; Singer et al., 2013). Nor should this be surprising, as we know from our own case running through scenarios helps anticipate how to act in similar situations. As for these sequences' generation, it is doubtful that they arise randomly, as Cheng et al. suggest of earlier observations. The fact that the location of the SWRs in Ólafsdóttir and colleagues' study could be directed by cues presented when they weren't in the maze suggests that their emergence isn't random or bound to a specific memory. Similarly, the idea that Mou and Ji's findings stem from random firing does little to account for the fact that they did not emerge in a bevy of control rats (including those seeing no track, those seeing and empty track, and those seeing a track with a toy car "running" it). More plausible, and what the authors themselves suggest, is that the rats are projecting as part of an overarching capacity for social learning. It remains to be seen just how creative rats can get—both studies involve environments that the rats had at least seen (a sequence unrelated to anything the animal has seen is obviously difficult to validate)—but we can say with some confidence that they go beyond their first-hand experience. This suffices for the basic claim—rats need not have the same horizons as humans to be engaged in the same process.

The case for extending inference beyond the spatial domain is more complex. Spatial tasks like maze running are the best-established paradigms for studying hippocampal inference, and it is a limitation of the present literature that non-spatial tasks are so much rarer (location being a difficult confound to control for). That said, few in the field expect planning and inference are purely spatial and for good reason (Redish, 2016; Eichenbaum, 2014; 2017a; Schiller et al., 2014). In the first case, there is evidence that the sequences observed during "spatial planning" aren't purely spatial. That is, they encode other, non-spatial information when relevant. For instance, one can tell based on the cells' firing rates when SWRs are guided by non-spatial cues of one or another type (e.g., one light or two), even when the path traced is the same (Takahashi, 2013; 2015). Second, there is no difference between spatial and non-spatial encoding that would prevent these mechanisms from being used in non-spatial domains. We know that hippocampal cells encode a great deal of non-spatial information in a manner that's essentially the same as the way they encode location and that these similarities extend to IGSs. Cells encoding temporal order, for instance, organize into theta-nested sequences just like location-encoding cells (Eichenbaum, 2014), while the tone maps reported by Aronov et al. (2017) house SWRs no different from their spatial counterparts. At the same time, we know that damage to the hippocampus undermines transitive and other forms of structural inference with stimuli like odors (Dusek and Eichenbaum, 1997; 1998) and touchscreen images (Silverman et al., 2015; Lazareva et al., 2015), suggesting that the mechanisms are not limited to space. Finally, it's worth noting that the characteristic back-and-forth looking behavior that’s been found to correlate with deliberation in spatial tasks has been observed in a variety of non-spatial tasks as well (Tolman, 1948; Hasz and Redish, 2018), consistent with the generalizability of VTE.

It’s possible, of course, that these various points aren’t related and that the spatial and non-spatial inferences interrupted by hippocampal damage have different underlying mechanisms. Direct recording evidence of IGSs being used in a contrastive, non-spatial decision-making task in a manner paralleling their operation in spatial tasks has yet emerge. Such a hypothesis would appear significantly less likely, however, as it would commit us to thinking that that the same populations, operating at the same frequencies, performing analogous inferential tasks, and associated with the same outward displays weren’t doing the same thing. At present, it’s unclear what this alternative mechanism would be. Positing such distinct mechanisms would also involve a high degree of evolutionary discontinuity and, frankly, a bit of a double standard. In humans, these inferential capacities have been bound together and linked to the hippocampus not by direct recording evidence but through coarse-grained techniques like functional imaging and lesion studies (Addis and Schacter, 2012; Kurth-Nelson et al., 2016). Requiring more fine-grained evidence to establish a shared neurobiological mechanism in rats would thus involve a degree of inconsistency. Finally, the best models available for spatial processes readily transfer to non-spatial domains (Gershman, 2017). State transition matrices and the learning algorithms that operate on them are, after all, indifferent to the kind of “states” involved. Locations work well, but there’s no reason why the same mechanism couldn’t operate on tones in a musical scale or sequences of odors (both of which are capable of being represented in rodent hippocampus). In effect, one would be positing additional mechanisms to save human uniqueness rather than account for the evidence. The better option in this case is simply to grant that the same mechanisms are at work.

4.5 Conclusion: Rational Animals

So, was Preston right? Were he and the Darwins, Tolmans, and Romaneses of comparative psychology correct to claim that no great gulf separates the working of our minds from that of other animals? If the evidence reviewed in the preceding pages holds, then I believe the answer must be yes—at least for beings with fronto-hippocampal networks comparable to our own. Reviewing the literature on rodents and primates, we find compelling evidence that humans are not, as Locke thought, the only species to abstract or form complex ideas. Rats, macaques, and a host of others appear capable of systematically deploying rules and concepts across a range of perceptually disparate conditions. Nor are we the only ones to anticipate future outcomes and possibilities, as Grotius contended. A Long-Evans Rat, it seems, is perfectly capable of putting itself in another's place or thinking about distant events. In fact, if what I have argued is correct, both capacities may be somewhat common. The evidence on which I have grounded my case comes mainly from rats and monkeys, groups relatively close to us genetically, but this is primarily because these are the animals for which relevant studies have been conducted. The underlying structures at issue are quite widespread, however, and what evidence is available suggests that they function in the same way. Bianchi's reports, for instance, indicate that frontal lobe damage causes comparable patterns of deficits in humans and dogs. This opens the door for far more species than previously suspected. If rats can think ahead, it suggests that animals with still more prominent frontal cortices and comparably organized hippocampal circuits may also. A list might include pigs, cats, horses, and—as may have pleased the King—dogs.

5.0 Postscript: Open Questions

It’s customary in concluding a work of this sort to consider open questions, and reviewing the foregoing chapters, there is no shortage of historical connections to unpack, empirical issues to parse, and lingering moral and conceptual points to forecast. Historically, more could and should be said of the animal psychology of the 1880s and 90s and of the debates surrounding the cognitive revolution, for instance. Of particular interest are the factors that helped re-establish the notion that minds could be measured, a point more or less assumed from the beginning of the third chapter but still controversial in Tolman’s time. Conceptually and empirically, the case for rationality would benefit from additional work on the process of belief revision, a dimension of rational activity bracketed here but subject to much study of late (Groman et al., 2019). Of all the issues left open, however, there are three that stand out to me: what the position defended here implies about our treatment of other animals, what it says about the origin and purpose of reasoning, and what, if not rationality, ultimately separates humans from other animals (why are humans so uniquely able to prove theorems, construct elaborate tales, and so forth). Each stand in need of far more intensive study than can be provided here, but the evidence reviewed does suggest a few lines of investigation. We can start with the moral question.

5.1 Moral Dimensions

As I noted in the introduction, a great part of the motivation behind the rationality debate stems from its ethical implications. The ability to plan, imagine, and re-live past experiences are commonly accorded substantial ethical import (Mendl and Paul, 2008; Suddendorf and Corballis, 2007). Such capacities are believed to enhance an agent's capacity for suffering and happiness, allowing as they do for experiences that stretch beyond the moment and comparisons that measure what is against what could be. Pains may be dreaded or relived, and their experience may take on a deeper significance when paired with an awareness of their alternatives. Beyond these experiential concerns, many ethicists and laypeople also believe that sophisticated cognitive capacities confer a special moral status on their holders (Sytsma and Machery, 2012). The intrinsic value of a life, it's argued, is enhanced by—if not dependent on—the value the organism is capable of attaching to it. Rocks and plants, it's supposed, lack any sense of themselves and consequently lack any interest in their continued existence. Human adults, by contrast, are keenly aware of what's going on in their lives and are interested in what the future may hold for them. Our future prospects matter not simply for others but for ourselves.

In part because of these implications, the capacity to imagine and weigh one's prospects has long been denied to animals like pigs, rats, and even other primates. The world seems a far more just place if, like Pope, we suppose that the lamb is incapable of grasping its impending slaughter (an argument made explicit by Malebranche, among others). Yet the results reviewed here, especially those regarding SWRs and VTE, provide strong evidence that this assumption is mistaken. The hippocampal formation does not cease to replay the past or trace out alternative scenarios when one moves beyond the human species. The dense recurrent connections that allow memories to be reconstructed and stitched together into sequences do not break down, nor do the sensory and emotional associations of these sequences go silent. To suppose, with full knowledge of the evidence reviewed and still accruing from studies of fear conditioning (e.g., Wu et al., 2017), that these neural processes are not accompanied by experience would, to quote Voltaire, mean entertaining the idea that “nature arranged all the means of feeling in [the] animal, so that it may not feel” (“Beasts”).

If one attaches ethical importance to decentered thought and if the results discussed earlier hold fast, however, many aspects of our treatment of other animals may need to be reconsidered. Beyond the immediate issue of commonplace killing or the infliction of physical pain, which has been widely discussed—and, I would argue, has a strong case against it solely on grounds of sentience—the capacity for reasoning presses concerns like living conditions. Animals in enclosed spaces, for instance, could suffer not only from the immediate distress of constraint but the peculiar boredom and sense of unfreedom that emerge when accompany the awareness of alternatives. Likewise, those deprived of a familiar conspecific or human (as happens when service animals retire), could suffer not only a sense of alarm in having their living situation changed but a feeling of loss over the familiar’s absence. If an animal is capable of simulating actions in environments it’s never visited (Mou and Ji, 2016) or replaying events in the service of a cognitive task (Singer et al., 2013), it’s at least plausible that they could ruminate over the loss of a pup (provided the species is one that forms such bonds). The evidence reviewed here does not tell us how these possibilities should be dealt with if they are present. We cannot yet say whether it would it be better to visit a retired service animal or if doing so would hinder them from starting a new life post-work, for instance. They do, however, lend an urgency to investigating such matters and recommend certain broad interventions (substantial cognitive enrichment of living spaces among them).[54] The assumption that other animals live in a constant present, visible in Pope’s *Essay on*

---

[54] Candidly, the research has led me to rethink the cognitive stimulation available in my cats’ living situation. I have taken to playing with them far more often as a result.

Man and popular writing like Robert Sapolsky's *Why Zebras Don't Get Ulcers*, is potentially quite dangerous.

Perhaps the greatest irony of the project is that the very studies that constitute my evidential base, studies I've argued are necessary to characterize animals' cognitive architectures with the confidence demanded by today's comparative psychologists, are not excepted from these implications. Invasive techniques like neural recording and lesioning obviously cause harm to the individual recorded or lesioned. To the extent that the results reviewed increase one's credence in the subjects' capacity to be harmed or up the moral significance of harms suffered, then, they weigh against the continued use of such techniques. More bluntly, I suspect the investigations are ethically self-undermining, establishing in their subjects the kind of mental lives that render such basic research immoral (I leave aside whether this applies for medical research). When combined with the arguments of chapter 3, this leaves the student of animal minds with a dilemma: either compromise on the ethics of invasive, curiosity-driven research or accept the uncertainty. Though I doubt it is the one animal researchers will take, my own sympathies lie with the latter. Many domains of study, as I mentioned in chapter three, have persistent evidential ambiguities. There are many domains of scholarship—history being one of them—that are accustomed to persistent interpretive disagreement and do not, in general, demand the kind of certainty-inducing, crucial evidence that comparative psychologists would like. This does not mean that these domains are less satisfying. Comparative cognition need not be one of these disciplines, I have argued, but if the alternative involves a serious moral compromise, the outcome doesn't look so bad. It may simply involve a return to the Darwinian way of doing things.

5.2 Evolution and Function

Transitioning from the ethical to the theoretical, a second question warranting further exploration concerns how we think about reasoning and its evolution. For years, architectures capable of abstraction and imaginative simulation have been assumed to involve "enormous expense of cognitive resources" relative to other designs (Suddendorf and Corballis, 2007, p. 307; Cf. Suddendorf et al., 2018). For this reason, their emergence has long been thought to require exceptional evolutionary circumstances, circumstances that came into being only recently. Ironically, the mental features and faculties that the Darwinians took, on evolutionary grounds, to be common between humans and a vast array of other mammals—sometimes even insects—have, in recent memory been thought to require pressures unique to early hominid life (see, e.g., Dennett, 1996). The results discussed in the final chapter suggest that this may not be the case.

To start with, we'll need a different time table. If the homology-based arguments are correct, then the latest we should date the "birth" of reason would be the date at which rodents and primates split. Going by common metrics, this takes us back around 87 million years (Springer et al., 2003). Estimates including carnivores like dogs and cats would need to go back 94 million years, while more liberal dates—including animals like elephants, say—would put the capacity back as far as 104 million years ago. In other words, reasoning is old. As for the evolutionary pressures responsible for its emergence, we can say that, whatever they were, they'd need to be something that acted on such a common ancestor. This leaves open more than a few candidates, but the leading pressures would seem to be more prosaic than the flashy capacities commonly cited in the literature. Flexible, explicit processes may be needed for language and other prototypically human activities, but they might be just as useful when it comes to remembering how to get to your nest or figuring out how to avoid a predator. Metabolically, it may be easier to maintain a single, multipurpose model-based architecture of the kind found in the hippocampus and PFC than to rely on a multiplicity of anatomically distinct and/or model-free architectures to perform the same range of tasks. In this sense, flexible and schematic cognitive architectures may simply have been the "cheaper" option.

Interestingly, recent developments in computational modelling and machine learning seem to point in the same direction. Over the last decade or so, flexible and schematic models have been developed that are biologically plausible, computationally feasible, and in line with the animal findings, exceedingly powerful when it comes to solving open-ended problems like spatial exploration and navigation (see, e.g., Schrittwieser et al., 2019). Any one task can be solved with architectures lacking the properties cited—another reason why finding an *experimentum crucis* in animal cognition is difficult—but solving a dozen or more distinct challenges with the same architecture becomes difficult. More precisely stated, architectures that directly map perceptual features onto specific actions or action-values (model-free architectures) struggle with generalization and adaptability in a way that architectures involving the construction of a formal environment model and the simulation of potential actions (model-based architectures) do not (Kaiser et al., 2020). Both types of architecture will do quite well when learning to play Atari games, for instance, but the ones that have the additional model-based component will be able to learn more games in the same amount of time and able to learn games with very different formats, such as Go and Chess (Schrittwieser et al., 2019). Combined with the biological data, these results seem to support the idea, propounded by Fodor (1987) among others, that the structured and systematic thought we think of as rational is visible even in commonplace behaviors. A cat getting its food in the morning might be just as reasoned an action as a Machiavellian primate thinking through how best to situate themselves in a social hierarchy.

5.3 Human Uniqueness

Finally, there is the question of what, if not the possession of imagination and rule-based cognitive control, makes us different. That humans are distinct from other animals with respect to certain intellectual achievements is undeniable. There is a reason why humans but no other species have gone to the moon, for instance. The question is why. Looking to the brain, there are a variety of features that one might point to (for a review, see Sousa et al., 2017). We undergo a considerably longer period of learning and development, have larger brains than our primate and rodent relatives, and manifest a brain-wide connectome that has adjusted to fit this increased size. We have larger and less excitable cortical neurons than our rodent relatives (Beaulieu-Laroche et al., 2018) as well as marked differences in astrocyte structure (structures that neonatal grafting studies have found to improve maze navigation and object memory in mice; Han et al., 2013). In terms of regional development, there are also a variety of anatomical and functional subdivisions that appear significantly enlarged in humans. The frontal pole, for instance, appears disproportionately large in humans relative to macaques and has been thought to underwrite our skills at multitasking (i.e., doing activities that require the tracking of multiple goals across changes in task; Mansouri et al., 2017).

All told, though, the differences can seem underwhelming. These shifts doubtlessly matter, but on their own, it’s hard to see how they could yield the behavioral divergence seen between humans and our closest relatives. They look, on the surface at least, more tweaks or additions than anything else. There’s a new layer here, a bit of growth there, but nothing that feels commensurate with the differences. The reason for this, I suspect, is that the processes that explain the most striking splits aren’t solely in the brain. When Huxley confronted the problem in *Man’s Place in Nature*, he symbolically located the difference between humans and animals in the hands. We are, he argued, quite similar in our mechanisms of thought but fundamentally different in our modes of interface with the world. I suspect this may be the bigger difference. As was discussed in the previous chapter, one of the principal neurobiological distinctions of humans is our motor system. We have a striking degree of direct cortical control over our movements served by, among other things, extensive corticomotoneuronal pathways to our hands and tongue. This is helpful not simply for things like tool-making but, when combined with neurobiological shifts related to language processing (Palomero-Gallagher et al., 2019), for interpersonal communication and coordination as well. We have the capacity not only to think complex thoughts but to express them through equally complex signals. These allow for the externalization of cognition and, over time, to its refinement and amplification. New generations pick up where the last left off and individuals learn, through a years-long enculturation process, to take up different sorts of rules and imagine scenarios of fundamentally new varieties. Were we limited to discovery by direct observation or individual insight, Huxley rightly observed, humanity likely never would have amassed the accomplishments on which it prides itself.

If this line of thought is correct, then the difference between human and animal is less about the "narrow" substrates of imaginative reasoning than the "extended" cognitive architecture of our environment and the sensorimotor structures that allow it to emerge (Clark, 2008; Sterelny, 2012). Small internal differences have an outsized effect when one considers the range of environmental interactions they allow—things like the calculation of lunar cycles, the writing of dialogues, and the undertaking of vast collaborative projects. In each case, the amount of the undertaking that we are able to keep in mind at once—even the ability to calculate large sums—is rather limited, as are the prospects of making serious progress in these projects without the social networks from which we acquire our language and practical skills. We progress on such matters by directing similar, if more powerful mechanisms toward challenges on which headway has already been made. The full story of human achievement, then, would seem to require looking beyond any one person to how concepts and problems change over long periods of time. One might look to the evolution of mathematical notation, say, or the development of novel physical concepts (for a quantitative look at collective problem-solving in Wikipedia edits, see DeDeo, 2013). Ultimately, though, this will require research of a different kind than we have been discussing. Rather than looking in the brain for the causes of the moon landing or the pyramids, I would suggest, we should be looking to history or sociology.

Bibliography

Abercrombie, J. (1838). *Inquiries Concerning the Intellectual Powers and the Investigation of Truth*. London: John Murray. Darwin's copy available at: https://www.biodiversitylibrary.org/bibliography/49184.

Alcaraz, F., Marchand, A., Courtand, G., Coutureau, E., & Wolff, M. (2016). "Parallel Inputs from the Mediodorsal Thalamus to the Prefrontal Cortex in the Rat." *European Journal of Neuroscience* 44 (3),1972–86. https://doi.org/10.1111/ejn.13316.

Allen, T. A., Salz, D., McKenzie, S., & Fortin, N. (2016). "Nonspatial Sequence Coding in CA1 Neurons." *Journal of Neuroscience* 36 (5),1547–1563.

Alvernhe, A., Save, E., & Poucet, B. (2011). Local remapping of place cell firing in the Tolman detour task. *European Journal of Neuroscience*, *33*(9), 1696–1705.

Alvernhe, A., VanCauter, T., Save, E., & Poucet, B. (2008). Different CA1 and CA3 representations of novel routes in a shortcut situation. *Journal of Neuroscience*, *28*(29), 7324–7333.

Amundson, R. (1983). EC Tolman and the intervening variable: A study in the epistemological history of psychology. *Philosophy of Science*, 50(2), 268–282.

Andrews, K., Comstock, G. L., G.K.D, C., Donaldson, S., Fenton, A., John, T. M., … Sebo, J. (2018). *Chimpanzee Rights: The Philosophers' Brief*. Routledge.

Anonymous. (n.d.). *Comment on Psychology's Loss of its Soul*. (James McKeen Cattell Papers, Box 80, Folder 3). Library of Congress, Washington D.C.

APC. (2002). Report of the Animal Procedures Committee 2002. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/251040/1128.pdf

Aquinas, T. (1990). *Summa Theologica*. Encyclopaedia Britannica.

Aronov, D., Nevers, R., & Tank, D. (2017). "Mapping of a Non-Spatial Dimension by the Hippocampal–Entorhinal Circuit." *Nature*, 543(7647):719–722.

Arroyo-Santos, A., Olson, M., & Vergara-Silva, F. (2014). The phylogeography debate and the epistemology of model-based evolutionary biology. *Biology & Philosophy*, *29*(6), 833–850.

“A Thought-Reading Dog”. (1903). *Animal Life and the World of Nature: a Magazine of Natural History Throughout the World, 1*, pp. 174-175. Retrieved from https://www.biodiversitylibrary.org/item/95282

Aust, U., Range, F., Steurer, M. & Huber, L. (2008). “Inferential Reasoning by Exclusion in Pigeons, Dogs, and Humans.” *Animal Cognition*, 11 (4):587–597.

Badre, D, and D’esposito, M. (2009). “Is the Rostro-Caudal Axis of the Frontal Lobe Hierarchical?” *Nature Reviews Neuroscience*, 10 (9):659–669.

Baldwin, J. M. (1913). *History of psychology: a sketch and an interpretation*. London: Watts & Co.

Balter, M. (2012). Animal cognition. “Killjoys” challenge claims of clever animals. *Science*, *335*(6072), 1036–1037.

Bartholomew, M. (1973). Lyell and Evolution: An Account of Lyell’s Response to the Prospect of an Evolutionary Ancestry for Man. *The British Journal for the History of Science*, *6*(3), 261–303.

Barton, R. A., & Venditti, C. (2014). Rapid Evolution of the Cerebellum in Humans and Other Great Apes. *Current Biology*, *24*(20), 2440–2444. https://doi.org/10.1016/j.cub.2014.08.056

Beadle, J. N., Tranel, D., Cohen, N. and Duff, M. (2013). “Empathy in Hippocampal Amnesia.” *Frontiers in Psychology*, 4, 1-12.

Beaulieu-Laroche, L., Toloza, E. H. S., Goes, M.-S. van der, Lafourcade, M., Barnagian, D., Williams, Z. M., Eskandar, E. N., Frosch, M. P., Cash, S. S., & Harnett, M. T. (2018). Enhanced Dendritic Compartmentalization in Human Cortical Neurons. *Cell*, *175*(3), 643-651.e14. https://doi.org/10.1016/j.cell.2018.08.045

Beard, G. (1882). *The Study of Trance, Muscle-reading and Allied Nervous Phenomena in Europe and America*. New York.

Bechtel, W. (2016). Investigating neural representations: The tale of place cells. *Synthese*, *193*(5), 1287–1321.

Beckers, T., De Houwer, J., & Dwyer, D. M. (2016). Reasoning versus association in animal cognition: Current controversies and possible ways forward. *Journal of Comparative Psychology*, *130*(3), 187–191. https://doi.org/10.1037/com0000024

Benhamou, S. (1996). No evidence for cognitive mapping in rats. *Animal Behaviour*, *52*(1), 201–212.

Bennett, A. (1996). Do animals have cognitive maps? *Journal of Experimental Biology*, *199*(1), 219–224.

Bennett, E. T. (1903). *The Society for Psychical Research: Its Rise & Progress & a Sketch of Its Work*. London: R. B. Johnson.

Bergmann, G. (1953). Theoretical Psychology. *Annual Review of Psychology*, *4*(1), 435–458.

Bianchi, F. M., Livi, L., Alippi, C., & Jenssen, R. (2017). Multiplex visibility graphs to investigate recurrent neural network dynamics. *Scientific Reports*, *7*, 44037. https://doi.org/10.1038/srep44037

Bianchi, L. (1922). *The mechanism of the brain: and the function of the frontal lobes*. E. & S. Livingstone.

Biddle, J. (2013). State of the field: Transient underdetermination and values in science. *Studies in History and Philosophy of Science Part A*, *44*, 124–133.

Bissonette, G. B., & Roesch, M. R. (2015). Neural correlates of rules and conflict in medial prefrontal cortex during decision and feedback epochs. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00266

Bissonette, G. B., & Roesch, M. R. (2017). Neurophysiology of rule switching in the corticostriatal circuit. *Neuroscience*, *345*, 64–76. https://doi.org/10.1016/j.neuroscience.2016.01.062

Binet, A. (1894). *The Psychic Life of Micro-organisms: A Study in Experimental Psychology*. Chicago: The Open Court Publishing Company.

Blackman, R. K., Crowe, D. A., DeNicola, A. L., Sakellaridi, S., MacDonald, A. W., & Chafee, M. V. (2016). Monkey prefrontal neurons reflect logical operations for cognitive control in a variant of the AX continuous performance task (AX-CPT). *Journal of Neuroscience*, *36*(14), 4067–4079.

Boakes, R. (1984). *From Darwin to Behaviourism: Psychology and the Minds of Animals*. Cambridge University Press.

Bokulich, A., & Oreskes, N. (2017). Models in Geosciences. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Science* (pp. 891–911). Springer International Publishing. https://doi.org/10.1007/978-3-319-30526-4_41

Bongard, S., & Nieder, A. (2010). Basic mathematical rules are encoded by primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, *107*(5), 2277–2282.

Boogert, N., Arbilly, M., Muth, F., & Seed, A. (2013). Do crows reason about causes or agents? The devil is in the controls. *Proceedings of the National Academy of Sciences*, 110(4), E273–E273.

Boring, E. G. (1950). *A History of Experimental Psychology*. Appleton-Century-Crofts.

Bostock, E., Muller, R., & Kubie, J. (1991). Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, *1*(2), 193–205.

Boswell, J., & Johnson, S. (1786). *Boswell's Life of Johnson: Including Boswell's Journal of a Tour of the Hebrides, and Johnson's Diary of A Journey Into North Wales*. Bigelow, Brown & Company, Incorporated.

Bozzano, E. (1905). Animals and psychic perception. *The Annals of Psychical Science*, *2*(7), 79-120.

Briskman, L. B. (1972). Is a Kuhnian analysis applicable to psychology? *Science Studies*, *2*(1), 87–97.

Browne, E. J. (1995). *Charles Darwin: Voyaging*. Knopf Doubleday Publishing Group.

Brunamonti, E., Mione, V., Di Bello, F., Pani, P., Genovesio, A., & Ferraina, S. (2016). Neuronal Modulation in the Prefrontal Cortex in a Transitive Inference Task: Evidence of Neuronal Correlates of Mental Schema Management. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(4), 1223–1236. https://doi.org/10.1523/JNEUROSCI.1473-15.2016

Buckner, C. (2014). The semantic problem(s) with research on animal mind-reading. *Mind & Language*, *29*(5), 566–589.

Buffon, G. L. L. (1797). *Buffon's Natural history. Containing a theory of the earth, a general history of man, of the brute creation, and of vegetables, minerals, &c. &c*. London: H. D. Symonds. http://archive.org/details/b28775892_0005

Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., & Miller, E. K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron*, *76*(4), 838–846.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Candland, D. K. (1995). Feral Children and Clever Animals: Reflections on Human Nature. Oxford University Press.

Carr, H. A. (1927). The interpretation of the animal mind. *Psychological Review*, *34*(2), 87–106. https://doi.org/10.1037/h0072244

Carr, H., & Watson, J. (1908). Orientation in the white rat. *Journal of Comparative Neurology and Psychology*, *18*(1), 27–44.

Carruthers, P. (1992). *The Animals Issue: Moral Theory in Practice*. Cambridge University Press.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.

Carruthers, P. (2013). Animal minds are real, (distinctively) human minds are not. *American Philosophical Quarterly*, *50*(3), 233–248.

Carruthers, P. (2015). *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. Oxford University Press.

Cattell, J. M. (1888). The psychological laboratory at Leipsic. *Mind*, *13*(49), 37–51.

Cattell, J. McKeen. (1898). Mrs. Piper, "The Medium." *Science, 7*(175), 640–42.

Chalmers, E., Luczak, A., & Gruber, A. J. (2016). Computational Properties of the Hippocampus Increase the Efficiency of Goal-Directed Foraging through Hierarchical Reinforcement Learning. *Frontiers in Computational Neuroscience*, *10*. https://doi.org/10.3389/fncom.2016.00128

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press.

Chang, L., Zhang, S., Poo, M., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences*, *114*(12), 3258–3263.

Chen, D., Lu, H., & Holyoak, K. J. (2014). The discovery and comparison of symbolic magnitudes. *Cognitive Psychology*, *71*, 27–54.

Chen, Z., Lowry, S., Jacobson, A., Hasselmo, M. E., & Milford, M. (2015). Bio-inspired homogeneous multi-scale place recognition. *Neural Networks*, *72*(Supplement C), 48–61. https://doi.org/10.1016/j.neunet.2015.10.002

Cheng, K., Shettleworth, S., Huttenlocher, J., & Rieser, J. (2007). Bayesian integration of spatial information. *Psychological Bulletin*, *133*(4), 625–637.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*(4), 391–416.

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews*, *60*, 82–89.

Chiappa, S., Racaniere, S., Wierstra, D., & Mohamed, S. (2017). Recurrent Environment Simulators. *ArXiv:1704.02254 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1704.02254

Choo, F., & Eliasmith, C. (2010). A Spiking Neuron Model of Serial-Order Recall. In *in 32nd Annual Conference of the Cognitive Science Society* (pp. 2188–2193).

Claparède, E. (1912). Les Chevaux Savants d'Elberfeld. *Archives de Psychologie*, *12*, 263–304.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

Clarke, A., & Fujimura, J. (2014). *The Right Tools for the Job: At Work in Twentieth-Century Life Sciences*. Princeton University Press.

Clifford, W. K. (1884). The Ethics of Belief. In *The Scientific Basis of Morals: And Other Essays* (pp. 25–35). New York: The Humboldt Publishing Company.

Cohn-Sheehy, B. I., & Ranganath, C. (2017). Time regained: how the human brain constructs memory for time. *Current Opinion in Behavioral Sciences, 17*, 169–177.

Coleridge, S. T. (1847). *Biographia Literaria Or Biographical Sketches of My Literary Life and Opinions by Tayl. Sam. Coleridge*. W. Pickering.

Coleridge, S. T. (1892). To a young ass. In *The Poetical Works of S.T. Coleridge* (p. 23). London: Frederick Warne and Co.

Colgin, L. L. (2016). Rhythms of the hippocampal network. *Nature Reviews Neuroscience*, *17*, 239–249 http://www.nature.com/nrn/journal/vaop/ncurrent/full/nrn.2016.21.html

Colgin, L., Moser, E., & Moser, M-B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences, 31*(9), 469–477.

Condillac, E. (2001). *Essay on the Origin of Human Knowledge*. Cambridge University Press.

Corballis, M. C. (2013). Mental time travel: a case for evolutionary continuity. *Trends in Cognitive Sciences*, *17*(1), 5–6. https://doi.org/10.1016/j.tics.2012.10.009

Corballis, M. C. (2017). *The Truth about Language: What it Is and where it Came from*. University of Chicago Press.

Cooke, W. (1877). *The Deity*. Hamilton, Adams, & Company.

Coon, D. J. (1992). Testing the limits of sense and science: American experimental psychologists combat spiritualism, 1880–1920. *American Psychologist*, *47*(2), 143–151.

Costall, A. (1993). How Lloyd Morgan's Canon Backfired. *Journal of the History of the Behavioural Sciences, 29*(2), 113–122.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron*, *66*(5), 796–807.

Cummings, A. M. (1888, June 30). Death Instinct in Animals. *Light*, p. 313. Retrieved from: http://www.iapsop.com/archive/materials/light/light_v8_jun_1888.pdf

Currie, A. (2018). *Rock, Bone, and Ruin: An Optimist's Guide to the Historical Sciences*. MIT Press.

Currie, A., & Levy, A. (2019). Why experiments matter. *Inquiry*, *62*(9–10), 1066–1090. https://doi.org/10.1080/0020174X.2018.1533883

Currie, A., & Turner, D. (2016). Introduction: Scientific Knowledge of the Deep Past. *Studies in History and Philosophy of Science Part A*, *55*, 43–46.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Darwin, C. (1873). *Origin of Certain Instincts*. *Nature*, *7*, 417–418. https://doi.org/10.1038/007417a0

Darwin, C. (1874). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray. 2d ed.

Darwin, C. (1887). *The Life and Letters of Charles Darwin, Including an Autobiographical Chapter*. Francis (ed.). London: John Murray.

Darwin, C. (1958). *Autobiography and Selected Letters*. (F. Darwin, Ed.). Courier Corporation.

Darwin, C. (2002). *The Complete Work of Charles Darwin Online* (http://darwin-online.org.uk/). John van Wyhe (ed.).

Darwin Correspondence Project, accessed February 2018, https://www.darwinproject.ac.uk/

Darwin, E. (1794a). *The Botanic Garden*. London: Printed for J. Johnson. Retrieved from http://archive.org/details/b28777554

Darwin, E. (1794b). *Zoonomia; Or, the Laws of Organic Life*. London: J. Johnson. Retrieved from http://archive.org/details/zoonomiaorlawsof1794darw

Darwin, E. (1803). *The Temple of Nature; or, The Origin of Society: A Poem, with Philosophical Notes*. London: J. Johnson. Retrieved from http://archive.org/details/templeofnatureor00darw

Daston, L., & Galison, P. (2007). *Objectivity*. Zone Books.

DeGrazia, D. (2002). *Animal Rights: A Very Short Introduction*. Oxford University Press.

DeDeo, S. (2013). Collective Phenomena and Non-Finite State Computation in a Human Social System. *PLOS ONE*, *8*(10), e75818. https://doi.org/10.1371/journal.pone.0075818

Del Arco, A., Park, J., Wood, J., Kim, Y., & Moghaddam, B. (2017). Adaptive encoding of outcome prediction by prefrontal cortex ensembles supports behavioral flexibility. *Journal of Neuroscience*, *37*(35), 8363–8373.

Dennett, D. C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. Basic Books.

deRochas, A. (1896). A dog’s laugh. *Popular Science Monthly*, *50*, 89–91.

“The Descent of Man: A Continuation of an Old Song” (1871). Blackwoods Magazine, 109(666), 517-9.

De Sio, F., & Marazia, C. (2014). Clever Hans and his effects: Karl Krall and the origins of experimental parapsychology in Germany. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 48, 94–102.

Desmond, A. (1989). *The Politics of Evolution*. Chicago: University of Chicago Press.

Desmond, A., & Moore, J. (1992). *Darwin: The Life of a Tormented Evolutionist*. New York: Grand Central Publications.

Desmond, A., & Moore, J. (2009). *Darwin’s Sacred Cause: How a Hatred of Slavery Shaped Darwin’s Views on Human Evolution*. Boston: Houghton Mifflin Harcourt.

Deuker, L., Bellmund, J. L., Navarro Schröder, T., & Doeller, C. F. (2016). An event map of memory space in the hippocampus. *ELife*, *5*. https://doi.org/10.7554/eLife.16534

DeVito, L. M., Lykken, C., Kanter, B. R., & Eichenbaum, H. (2010). Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. *Learning & Memory*, *17*(3), 161–167.

Diester, I., & Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biology*, *5*(11), e294.

Duncan, J., & Miller, E. K. (2002). Cognitive focus through adaptive neural coding in the primate prefrontal cortex. *Principles of Frontal Lobe Function*, 278–291.

Dunlap, K. (1922). *The Elements of Scientific Psychology*. C.V. Mosby Company.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences*, *94*(13), 7109–7114.

Dusek, J. A., & Eichenbaum, H. (1998). The hippocampus and transverse patterning guided by olfactory cues. *Behavioral Neuroscience*, *112*(4), 762-771.

Duverne, S., & Koechlin, E. (2017). Hierarchical Control of Behaviour in Human Prefrontal Cortex. *The Wiley Handbook of Cognitive Control*, 207–220.

Eichenbaum, H. (2017a). On the Integration of Space, Time, and Memory. *Neuron; Cambridge*, *95*(5), 1007–1018.

Eichenbaum, H. (2017b). Prefrontal-hippocampal interactions in episodic memory. *Nature Reviews Neuroscience*, *18*(9), 547–558.

Eichenbaum, H. (2017c). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, *117*(4), 1785–1796.

Eichenbaum, H., … Witter, M. (2016). Hippocampus at 25. *Hippocampus*, 26(10), 1238–1249.

Eiselt, A.-K., & Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. *Journal of Neuroscience*, *33*(17), 7526–7534.

Eliasmith, C. (2005). Cognition with neurons: A large-scale, biologically realistic model of the Wason task. In B. Bara, L. Barasalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 624– 629). Mahwah, NJ: Lawrence Erlbaum Associates.

Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, *338*(6111), 1202–1205.

Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives on Psychological Science*, *13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Erdohegyi, Á., Topál, J., Virányi, Z., & Miklósi, Á. (2007). Dog-logic: inferential reasoning in a two-way choice task and its restricted use. *Animal Behaviour*, *74*(4), 725–737.

Evans, E. (1893). Studies of animal speech. *Popular Science Monthly*, *43*, 433–439.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Fairhall, A. (2014). The receptive field is dead. Long live the receptive field? *Current Opinion in Neurobiology*, *25*, ix–xii. https://doi.org/10.1016/j.conb.2014.02.001

Farquhar, M., & Palade, G. (1981). The Golgi apparatus (complex)-(1954-1981)-from artifact to center stage. *Journal of Cell Biology*, *91*(3), 77–103.

Fauconnier, G., & Turner, M. (2003). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

Fearon, D. R. (1901). Histrionic capacity of Grey Parrots. *Nature*, *65*(1676), 127. https://doi.org/10.1038/065127a0

Fitt, H. (1890). *The Scientific Education of Dogs for the Gun*. London: Sampson, Low & Co.

Fitzpatrick, S., & Goodrich, G. (2017). Building a Science of Animal Minds: Lloyd Morgan, Experimentation, and Morgan's Canon. Journal of the History of Biology, 50(3), 525–569. https://doi.org/10.1007/s10739-016-9451-x

Floridi, L. (1997). Scepticism and Animal Rationality: The Fortune of Chrysippus' Dog in the History of Western Thought. *Archiv Für Geschichte Der Philosophie*, *79*(1), 27–57. https://doi.org/10.1515/agph.1997.79.1.27

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. The MIT Press.

Foster, D. J. (2017). Replay Comes of Age. *Annual Review of Neuroscience*, *40*(1), 581–602. https://doi.org/10.1146/annurev-neuro-072116-031538

Fotherby, H. (1908). Music and emotion. *The Annals of Psychical Science*, 7(42), 261–309.

Franklin, A. (1989) The epistemology of experiment. In Gooding, D., Pinch, T., & Schaffer, S. (eds). *The Uses of Experiment: Studies in the Natural Sciences*. Cambridge University Press.

Franklin, L. (2005). Exploratory Experiments. *Philosophy of Science*, 72(5), 888–899.

Frankland, S. M., & Greene, J. D. (2020). Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annual Review of Psychology*, *71*(1), 273–303. https://doi.org/10.1146/annurev-psych-122216-011829

Frigerio, A., Giordani, A., & Mari, L. (2010). Outline of a general model of measurement. *Synthese*, *175*(2), 123–149.

Fudge, E. (2005). *Brutal reasoning: animals, rationality, and humanity in early modern England*. Cornell University Press.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*, 66–74.

Fuster, J. (2015). *The Prefrontal Cortex*. Academic Press.

Fyfe, A. (1997). The reception of William Paley's *Natural Theology* in the University of Cambridge. *The British Journal for the History of Science*, *30*(3), 321–335. https://doi.org/10.1017/S0007087497003117

G. H. P. (1899). Animal hypnotism. *The American Naturalist*, *33*(389), 434–436. https://doi.org/10.1086/277279

Galef, B. G. (1998). Edward Thorndike: Revolutionary psychologist, ambiguous biologist. *American Psychologist*, *53*(10), 1128-1134.

Galloway, W. (1906). The Intelligence of Animals. *Nature*, *73*(1897), 440. https://doi.org/10.1038/073440a0

Gallup, G. G. (1970). Chimpanzees: self-recognition. *Science*, *167*(3914), 86–87.

Gates, E. (1895). The science of mentation and some new general methods of psychologic research. *The Monist*, *5*(4), 574–597.

Gauthier, J. L., & Tank, D. W. (2017). Context-invariant encoding of reward location in a distinct hippocampal population. *BioRxiv*, 207043.

Gentry, G., Brown, W., & Kaplan, S. (1947). An experimental analysis of the spatial location hypothesis in learning. *Journal of Comparative and Physiological Psychology*, *40*(5), 309.

Gershman, S. J. (2017). Predicting the Past, Remembering the Future. *Current Opinion in Behavioral Sciences*, *17*, 7–13. https://doi.org/10.1016/j.cobeha.2017.05.025

Gibson, S. (2015). *Animal, Vegetable, Mineral: How eighteenth-century Science Disrupted the Natural Order*. Oxford University Press.

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*(8), 618–631.

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*(10), 1222–1223. https://doi.org/10.1038/nn.2384

Godwin, W. (1793). *An Enquiry Concerning Political Justice, and its Influence on General Virtue and Happiness*. London: G.G.J. and J. Robinson

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485.

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

Goodroe, S., Starnes, J., & Brown, T. (2018). The Complex Nature of Hippocampal-Striatal Interactions in Spatial Navigation. *Frontiers in Human Neuroscience*, *12*.

Greenwood, J. D. (2016). All the way up or all the way down?: Some historical reflections on theories of psychological continuity. *Journal of Comparative Psychology*, *130*(3), 205-214.

Groman, S. M., Keistler, C., Keip, A. J., Hammarlund, E., DiLeone, R. J., Pittenger, C., Lee, D., & Taylor, J. R. (2019). Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes. *Neuron*, *103*(4), 734-746.e3. https://doi.org/10.1016/j.neuron.2019.05.042

Groos, K. (1898). *The Play of Animals*. New York: D. Appleton.

Grotius, H. (1738). *The Rights of War and Peace, in Three Books: Wherein are Explained, the Law of Nature and Nations, and the Principal Points Relating to Government*. The Lawbook Exchange, Ltd.

Gruber, H. E., & Barrett, P. (1974). *Darwin: A Man, a Study of Scientific Creativity*. Wildwood House.

Güntürkün, O., & Bugnyar, T. (2016). Cognition without cortex. *Trends in Cognitive Sciences*, *20*(4), 291–303.

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2012). Segmentation of spatial experience by hippocampal θ sequences. *Nature Neuroscience*, *15*(7), 1032–1039. https://doi.org/10.1038/nn.3138

Gupta, A. S., van der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, *65*(5), 695–705.

Gurney, E., Myers, F., & Podmore, F. (1886). *Phantasms of the Living*, Volume I. London: Rooms of the Society for Psychical Research, Trübner and Company.

Guthrie, E. R. (1935). *Psychology of Learning.* New York: Harper.

Hacking, I. (1983). *Representing and intervening*. Cambridge University Press.

Hakeem, A. Y., Sherwood, C. C., Bonar, C. J., Butti, C., Hof, P. R., & Allman, J. M. (2009). Von Economo Neurons in the Elephant Brain. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, *292*(2), 242–248. https://doi.org/10.1002/ar.20829

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–831.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505.

Halina, M. (2015). There is no special problem of mindreading in nonhuman animals. *Philosophy of Science*, *82*(3), 473–490.

Han, X., Chen, M., Wang, F., Windrem, M., Wang, S., Shanz, S., Xu, Q., Oberheim, N. A., Bekar, L., & Betstadt, S. (2013). Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell Stem Cell*, *12*(3), 342–353.

Hanus, D. (2016). Causal reasoning versus associative learning: A useful dichotomy or a strawman battle in comparative psychology? *Journal of Comparative Psychology*, *130*(3), 241-248.

Harrison, P. (1993). Animal Souls, Metempsychosis, and Theodicy in Seventeenth-Century English Thought. *Journal of the History of Philosophy*, *31*(4), 519–544. https://doi.org/10.1353/hph.1993.0081

Haselgrove, M. (2016). Overcoming associative learning. *Journal of Comparative Psychology*, *130*(3), 226-240.

Hartley, D. (1749). *Observations on Man, his Frame, his Duty, and his Expectations*. London: S.D. Richards.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726–1731.

Hasz, B. M., & Redish, A. D. (2018). Deliberation and Procedural Automation on a Two-Step Task for Rats. *Frontiers in Integrative Neuroscience*, *12*. https://doi.org/10.3389/fnint.2018.00030

Henley, T. (2017). Hergenhahn's An Introduction to the History of Psychology. Cengage Learning.

Henshaw, S. (1897). Proceedings of Scientific Societies. The American Naturalist, 31(366), 559–566.

Herold, C., Palomero-Gallagher, N., Hellmann, B., Kröner, S., Theiss, C., Güntürkün, O., & Zilles, K. (2011). The receptor architecture of the pigeons' nidopallium caudolaterale: an avian analogue to the mammalian prefrontal cortex. *Brain Structure and Function*, *216*(3), 239–254.

Herrick, C. L. (1895). Popular Zoological literature. *The Journal of Comparative Neurology*, *5*, i-iv.

Herrick, F. (1910). Instinct and Intelligence in Birds. *Popular Science Monthly*, *76*, 532–556.

Hesslow, G. (2012). The current status of the simulation theory of cognition. *Brain Research*, *1428*, 71–79.

Heyes, C. (2012). Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Heyes, C. (2015). Animal mindreading: what's the problem? *Psychonomic Bulletin & Review*, *22*(2), 313–327. https://doi.org/10.3758/s13423-014-0704-4

Heyes, C. (2017). Apes submentalise. *Trends in Cognitive Sciences*, *21*(1), 1–2.

Hill, C., & Thune, L. (1952). Place and response learning in the white rat under simplified and mutually isolated conditions. *Journal of Experimental Psychology*, 43(4), 289-297.

Hobhouse, L. (1901). Mind in Evolution. New York: Macmillan.

Holmes, S. J. (1916). *Studies in Animal Behavior*. Boston, R. G. Badger.

Holt, E. B. (1915). Response and Cognition I: The Specific-Response Relation. *The Journal of Philosophy, Psychology and Scientific Methods*, *12*(14), 365–373. https://doi.org/10.2307/2013669

Hudson, T. J. (1893). *The Law of Psychic Phenomena: A Working Hypothesis for the Systematic Study of Hypnotism, Spiritism, Mental Therapeutics, Etc*. A. C. McClurg and Company.

Hughes, D. J. (2006). Proofs without syntax. *Annals of Mathematics*, *164*, 1065–1076.

Hunt, L. T., & Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience*, *18*(3), 172–182.

Hurley, S. L., & Nudds, M. (2006). *Rational Animals?* Oxford University Press.

Hussar, C. R., & Pasternak, T. (2013). Common rules guide comparisons of speed and direction of motion in the dorsolateral prefrontal cortex. *Journal of Neuroscience*, *33*(3), 972–986.

Huttunen, A. W., Adams, G. K., & Platt, M. L. (2017). Can self-awareness be taught? Monkeys pass the mirror test—again. *Proceedings of the National Academy of Sciences*, *114*(13), 3281–3283.

Huxley, T. H. (1886). *Hume*. London: Macmillan. Retrieved from http://archive.org/details/humeletters00huxluoft

Isaac, J., Buchanan, K., Muller, R., & Mellor, J. (2009). Hippocampal Place Cell Firing Patterns Can Induce Long-Term Synaptic Plasticity In Vitro. *Journal of Neuroscience*, *29*(21), 6840–6850.

Innis, N. (1998) History of comparative psychology in biographical sketches. In G. Greenberg, and M. Haraway (Eds.). *Comparative Psychology: A Handbook* (pp. 3-24). Routledge.

Jafarpour, A., Piai, V., Lin, J. J., & Knight, R. T. (2017). Human hippocampal pre-activation predicts behavior. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-06477-5

James, W. (1886). Letter to Stumpf. In H. James (Ed.). *The Letters of William James*. (Vol. I). Boston: Atlantic Monthly Press.

James, W. (1890a). *The Principles of Psychology Volume II*. Henry Holt & Co.

James, W. (1890b). Report on Experiments with Mrs. Piper. *Proceedings of the Society for Psychical Research, 6*, 651-659.

James, William (1896). The Will to Believe. *The New World*, *5*, 327–347.

James, William (1898). Review of R. Hodgson's "A further record of observations of certain phenomena of trance". *Psychological Review*, *5*, 420–424.

James, William (1986). What psychical research has accomplished. In Burkhardt, F., Bowers, F., & Skrupskelis, I. (Eds.), *The Works of William James, Volume 16: Essays in Psychical Research*. Harvard University Press.

Jastrow, J. (undated September 16th). Letter on pummeling mysticism. (James McKeen Cattell Papers, box 24, Jastrow, J.). Library of Congress, Washington D.C.

Jastrow, Joseph. (1889). The problems of "psychic research." *Harper's New Monthly Magazine*, *79*(469), 76.

Jastrow, J. (1900 November 7). *Letter to Cattell*. (James McKeen Cattell Papers, box 24, Jastrow, J.). Library of Congress, Washington D.C.

Jastrow, Joseph. (1900). *Fact and Fable in Psychology*. New York: Houghton Mifflin.

Jastrow, J. (1901). Some currents and undercurrents in psychology. Psychological Review, 8(1), 1–26.

Jastrow, J. (1906). Fact and fable in animal psychology. *Popular Science Monthly*, *69*, 138–146.

Jastrow, Joseph. (1930). *Wish and Wisdom: Episodes in the Vagaries of Belief*. D. Appleton-Century Company.

Jastrow, Joseph. (1935). Joseph Jastrow. In C. Murchison (Ed.), *A History of Psychology in Autobiography* (Vol. I, pp. 135–62). Worcester, MA: Clark University Press.

J. E. A. T. (1904). Intelligence in animals. *Nature*, *71*(1831), 102.

Jennings, H. S. (1908a). Recent work on the behavior of higher animals. *The American Naturalist*, *42*(495), 207–216. https://doi.org/10.1086/278924

Jennings, H. S. (1908b). Recent work on the behavior of higher animals. *The American Naturalist*, *42*(497), 355–360. https://doi.org/10.1086/278943

Jennings, H. S. (1908c). Mind in animals. *The American Naturalist*, *42*(503), 754–760. https://doi.org/10.1086/279007

Jennings, H. S. (1913). Causes and determiners in radically experimental analysis. *The American Naturalist*, *47*(558), 349–360. https://doi.org/10.1086/279355

Jelsing, J., Hay-Schmidt, A., Dyrby, T., Hemmingsen, R., Uylings, H. B. M., & Pakkenberg, B. (2006). The prefrontal cortex in the Göttingen minipig brain defined by neural projection criteria and cytoarchitecture. *Brain Research Bulletin*, *70*(4–6), 322–336.

Jensen, G., Altschul, D., Danly, E., & Terrace, H. (2013). Transfer of a serial representation between two distinct tasks by rhesus macaques. *PLoS One*, *8*(7), e70285.

Jin, J., Chen, Q., Qiao, Q., Yang, L., Xiong, J., Xia, J., … Chen, F. (2016). Orexin neurons in the lateral hypothalamus project to the medial prefrontal cortex with a rostro-caudal gradient. *Neuroscience Letters*, *621*, 9–14.

Johnson, A., & Crowe, D. A. (2009). Revisiting Tolman, his theories and cognitive maps. *Cognitive Critique*, *1*, 43–72.

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience*, *27*(45), 12176–12189. Retrieved from http://www.jneurosci.org/content/27/45/12176.short.

Johnson-Laird, Philip N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250.

Johnson-Laird, Philip Nicholas. (2006). *How We Reason*. Oxford University Press.

Jonçich, G. M. (1968). *The Sane Positivist: A Biography of Edward L. Thorndike*. Wesleyan University Press.

Jones, J. (1861). *The Natural and Supernatural; or Man Physical, Apparitional, and Spiritual*. London: H. Bailliere.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahnemann, P. Slovic & A.Tversky (Eds.)*: Judgement Under Uncertainty: Heuristics and biases.*(pp. 202-8) Cambridge University Press.

Kaiser, L., Babaeizadeh, M., Milos, P., Osinski, B., Campbell, R. H., Czechowski, K., Erhan, D., Finn, C., Kozakowski, P., Levine, S., Mohiuddin, A., Sepassi, R., Tucker, G., & Michalewski, H. (2020). Model-Based Reinforcement Learning for Atari. *ArXiv:1903.00374 [Cs, Stat]*. http://arxiv.org/abs/1903.00374

Kamiński, J., Sullivan, S., Chung, J. M., Ross, I. B., Mamelak, A. N., & Rutishauser, U. (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nature Neuroscience*, *20*(4), 590–601.

Kendler, H. (1947). An investigation of latent learning in a T-maze. *Journal of Comparative and Physiological Psychology*, 40(4), 265-270.

Klein, G. (1998). *Sources of power: how people make decisions*. MIT Press.

Kleinhans, M. G., Buskes, C. J., & de Regt, H. W. (2005). Terra Incognita: explanation and reduction in earth science. *International Studies in the Philosophy of Science*, *19*(3), 289–317.

Kline, L. W. (1899). Methods in animal psychology. *The American Journal of Psychology*, *10*(2), 256–279.

Knowlton, B. J., Morrison, R. G., Hummel, J. E., & Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends in Cognitive Sciences*, *16*(7), 373–381.

Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Current Opinion in Neurobiology*, *37*, 1–6.

Kohda, M., … Jordan, A. (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLoS Biology*, *17*(2), e3000021.

Korsgaard, C. (2004). Fellow creatures: Kantian ethics and our duties to animals. In G.B. Peterson (ed.) *The Tanner Lectures on Human Values*. Salt Lake City: University of Utah Press, pp. 79–110.

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Harvard University Press.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97–109. https://doi.org/10.1037/a0020762

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, *20*(7), 512–534.

Lacassagne, A. (1882). Criminality in animals. In *Popular Science Monthly*, 22, 244-255.

La Mettrie, J. O. de. (1912). *Man a Machine*. Chicago: Open Court Publishing. Retrieved from http://archive.org/details/manmachine00lame

Lamont, P. (2012). The making of extraordinary psychological phenomena. *Journal of the History of the Behavioral Sciences*, *48*(1), 1–15. https://doi.org/10.1002/jhbs.21516

Lara, A. H., & Wallis, J. D. (2014). Executive control processes underlying multi-item working memory. *Nature Neuroscience*, *17*(6), 876–883. https://doi.org/10.1038/nn.3702

Lazareva, O. (2012). Transitive inference in nonhuman animals. In E. Wasserman & T. Zentall (Eds.) *The Oxford Handbook of Comparative Cognition* (pp. 718-735). Oxford University Press.

Lazareva, O. F., Kandray, K., & Acerbo, M. J. (2015). Hippocampal lesion and transitive inference: Dissociation of inference-based and reinforcement-based strategies in pigeons. *Hippocampus*, *25*(2), 219–226. https://doi.org/10.1002/hipo.22366

Leahey, T. H., & Leahey, G. E. (1983). *Psychology's Occult Doubles: Psychology and the Problem of Pseudoscience*. Chicago: Nelson-Hall.

Leavens, D., Bard, K., & Hopkins, W. (2019). The mismeasure of ape social cognition. *Animal Cognition*, *22*(4), 1–18.

Lemon, R. (2008). Descending pathways in motor control. *Annual Review of Neuroscience*, *31*, 195–218.

Letourneau, C. (1898). Education in the animal kingdom. *Popular Science Monthly*, *52,* 527–533.

Lever, C., Kaplan, R., & Burgess, N. (2014). The Function of Oscillations in the Hippocampal Formation. In D. Derdikman & J. Knierim (Eds.), *Space, Time and Memory in the Hippocampal Formation* (pp. 303–350). Vienna: Springer Vienna.

Lewes, G. H. (1879). *Problems of Life and Mind*. London: Trübner.

Lindemann-biolsi, K. L., & Reichmuth, C. (2014). Cross-modal transitivity in a California sea lion (Zalophus californianus). *Animal Cognition; Heidelberg*, *17*(4), 879–890.

Lindsay, W. L. (1880). Mind in the lower animals, in health and disease. New York: D. Appleton and Company.

Lisman, J., Buzsáki, G., Eichenbaum, H., Nadel, L., Rangananth, C., & Redish, A. D. (2017). Viewpoints: how the hippocampus contributes to memory, navigation and cognition. *Nature Neuroscience*, *20*(11), 1434-1447

Livingstone, M. S., Pettine, W. W., Srihasam, K., Moore, B., Morocz, I. A., & Lee, D. (2014). Symbol addition by monkeys provides evidence for normalized quantity coding. *Proceedings of the National Academy of Sciences*, *111*(18), 6822–6827.

Locke, J. (1847). *An Essay Concerning Human Understanding*. Philadelphia: Kay & Troutman.

Long, W. J. (1904). Science, nature and criticism. *Science, 19*(489), 760-767.

Lovejoy, A. O. (1936). *The Great Chain of Being: A Study of the History of an Idea*. Harvard University Press.

Lubbock, S. J. (1877). On the habits of ants. In *Popular Science Monthly*, 11, 39-58.

Ludy, B. (2006). *A History of Psychology in Letters*. New York: Blackwell.

MacDougall, R. (1899a). The dawn of reason. *The American Naturalist*, *33*(391), 611–613. https://doi.org/10.1086/277366

MacDougall, R. (1899b). The methods of comparative psychology. *The American Naturalist*, *33*(391), 613–615. https://doi.org/10.1086/277367

Mackintosh (sir.), J. (1835). *Memoirs of the Life of Sir James Mackintosh* R.J. Mackintosh (Ed.). London: E. Moxon.

Maguire, R. (1907). The study of animals. *The Theosophical Review*, 39(234), 535-537.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., & Robbins, T. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, *125*(3), 624–639. https://doi.org/10awf049.pdf3/brain/awf049

Mansouri, F. A., Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing competing goals—A key role for the frontopolar cortex. *Nature Reviews Neuroscience*, *18*(11), 645–657. https://doi.org/10.1038/nrn.2017.111

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617. https://doi.org/10.1038/s41593-018-0232-z

Mayor, J. E. (1898). King James I. on the reasoning faculty in dogs. *Classical Review*, *12*, 93–96.

McEachran, D. (1888). *Opening address Society of Comparative Psychology* Montreal: Gazette Publishing Company. Retrieved from http://archive.org/details/cihm _09413

McGetrick, J., & Range, F. (2018). Inequity aversion in dogs: a review. *Learning & Behavior*, *46*(4), 479–500.

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, *83*(1), 202–215.

McKenzie, S., Keene, C. S., Farovik, A., Bladon, J., Place, R., Komorowski, R., & Eichenbaum, H. (2016). Representation of memories in the cortical-hippocampal system: Results from the application of population similarity analyses. *Neurobiology of Learning and Memory*, *134 Pt A*, 178–191. https://doi.org/10.1016/j.nlm.2015.12.008

Mendl, M., & Paul, E. S. (2008). Do animals live in the present?: Current evidence and implications for welfare. *Applied Animal Behaviour Science*, *113*(4), 357–382.

Mian, M. K., Sheth, S. A., Patel, S. R., Spiliopoulos, K., Eskandar, E. N., & Williams, Z. M. (2014). Encoding of rules by neurons in the human dorsolateral prefrontal cortex. *Cerebral Cortex*, *24*(3), 807–816. https://doi.org/10.1093/cercor/bhs361

Milivojevic, B., Varadinov, M., Grabovetsky, A. V., Collin, S. H., & Doeller, C. F. (2016). Coding of event nodes and narrative context in the hippocampus. *Journal of Neuroscience*, *36*(49), 12412–12424.

Milivojevic, B., Vicente-Grabovetsky, A., & Doeller, C. F. (2015). Insight reconfigures hippocampal-prefrontal memories. *Current Biology*, *25*(7), 821–830.

Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., Hefft, S., Merkow, M., Polyn, S. M., Jacobs, J., & others. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science*, *342*(6162), 1111–1114.

Millikan, Ruth G. (2006). Styles of Rationality. In S. L. Hurley & M. Nudds (Eds.), *Rational Animals?* Oxford University Press.

Millikan, Ruth Garrett. (2004). *Varieties of Meaning*. MIT press.

Mills, W. (1897). Psychology and comparative psychology. *Science*, *5*(123), 718–720.

Mills, W. (1898). *The Nature and Development of Animal Intelligence*. New York: Macmillan.

Mills, W. (1904). Some Aspects of the Development of Comparative Psychology. *Science*, *19*(489), 745-757.

Mininni, C. J., & Zanutto, B. S. (2017). Exploring the limits of learning: Segregation of information integration and response selection is required for learning a serial reversal task. *PloS One*, *12*(10), e0186959.

M. N. W. (1912). Animal intelligence. *Nature*, *89*(2217), 192. https://doi.org/10.1038/089192b0

Momennejad, I. (2020). *Learning Structures: Predictive Representations, Replay, and Generalization* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/b6sr8

Monboddo, J. B. (1774). *Of the Origin and Progress of Language*. Edinburgh: J. Balfour. Retrieved from http://archive.org/details/originandprogre06monbgoog

Monti, M. M., & Osherson, D. N. (2012). Logic, language and the brain. *Brain Research*, *1428*, 33–42. https://doi.org/10.1016/j.brainres.2011.05.061

Moore, R. L. (1977). *In Search of White Crows: Spiritualism, Parapsychology, and American Culture*. Oxford University Press.

Morgan, C. L. (1894). *An Introduction to Comparative Psychology*. W. Scott, limited.

Morris, R. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, *12*(2), 239–260.

Moses, S. N., & Ryan, J. D. (2006). A comparison and evaluation of the predictions of relational and conjunctive accounts of hippocampal function. *Hippocampus*, *16*(1), 43–65.

Moskaleva, M., & Nieder, A. (2014). Stable numerosity representations irrespective of magnitude context in macaque prefrontal cortex. *European Journal of Neuroscience*, *39*(5), 866–874.

Mou, X., & Ji, D. (2016). Social observation enhances cross-environment activation of hippocampal place cell patterns. *ELife*, *5*, e18022. https://doi.org/10.7554/eLife.18022

Müller, F. M. (Ed.). (1879). *The Upanishads* (Vol. 1). Clarendon Press.

Müller, F. M. (1887). *The Science of Thought*. New York: Charles Scribner's Sons.

Münsterberg, H. (1889). Gedankenübertragung. Vortrag gehalten in der Akademischen Gesellschaft zu Freiburg i. B. am 10. Januar 1889. Freiburg, Germany: Mohr.

Münsterberg, H. (1899). Psychology and mysticism. *Atlantic Monthly*, *LXXXIII*, 67–85.

Münsterberg, H. (1913 January 29). *Letter to Cattell*. (James McKeen Cattell Papers, box 136, Münsterberg, H.). Library of Congress, Washington D.C.

Münsterberg, H. (1914). *Psychology and Social Sanity*. Garden City, NY: Doubleday, Page & Company.

Münsterberg, H. (1915). *Harvard Psychological Studies* (Vol. IV). Harvard University Press.

Myers, F. (1888). Proceedings of the Society for Psychical Research. *Science*, *ns-12*(292), 118–118. https://doi.org/10.1126/science.ns-12.292.118

Myers, F. (1893). Prof. Wundt on hypnotism and suggestion. *Mind*, *2*(5), 95–101. https://doi.org/10.1093/mind/II.5.95

Myers, F. and Sully, J. (1892). The International Congress of Experimental Psychology. *Nature*, *46*, 261-262. https://doi.org/10.1038/046362b0

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*, *109*(29), 11860–11865.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382.

Nieder, A., & Merten, K. (2007). A Labeled-Line Code for Small and Large Numerosities in the Monkey Prefrontal Cortex. *Journal of Neuroscience*, *27*(22), 5986–5993. https://doi.org/10.1523/JNEUROSCI.1056-07.2007

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences*, *112*(35), 11078–11083. https://doi.org/10.1073/pnas.1507104112

O'Donnell, J. (1985). The origins of behaviorism: American psychology, 1870-1920. New York: New York University Press.

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Clarendon Press.

Okuyama, T., Kitamura, T., Roy, D. S., Itohara, S., & Tonegawa, S. (2016). Ventral CA1 neurons store social memory. *Science*, *353*(6307), 1536–1541.

Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D., & Spiers, H. J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *ELife*, *4*, e06063. https://doi.org/10.7554/eLife.06063

Oorschot, D. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: a stereological study using the cavalieri and optical disector methods. *Journal of Comparative Neurology*, 366(4), 580-599.

Oppenheim, J. (1988). *The Other World: Spiritualism and Psychical Research in England, 1850-1914*. Cambridge University Press.

Osvath, M. (2016). Putting flexible animal prospection into context: escaping the theoretical box. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(1), 5–18.

Osvath, M., & Martin-Ordas, G. (2014). The future of future-oriented cognition in non-humans: theory and the empirical case of the great apes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655). https://doi.org/10.1098/rstb.2013.0486

Packard, M. (2009). Anxiety, cognition, and habit: a multiple memory systems perspective. *Brain Research*, 1293, 121–128.

Packard, M. G., & Goodman, J. (2013). Factors that influence the relative use of multiple memory systems. *Hippocampus*, *23*(11), 1044–1052. https://doi.org/10.1002/hipo.22178

Palomero-Gallagher, N., & Zilles, K. (2019). Differences in cytoarchitecture of Broca's region between human, ape and macaque brains. *Cortex*, *118*, 132–153. https://doi.org/10.1016/j.cortex.2018.09.008

Pannese, E. (1999). The Golgi Stain: Invention, Diffusion and Impact on Neurosciences. *Journal of the History of the Neurosciences*, *8*(2), 132–140.

Peckham G. W., & Peckham, E. G. (1887). Some observations on the mental powers of spiders. *Journal of Morphology*, *1*(2), 383–419. https://doi.org/10.1002/jmor.1050010206

Penn, D. (2011). How folk psychology ruined comparative psychology: and how scrub jays can save it. *Animal Thinking: Contemporary Issues in Comparative Cognition* (pp. 253–266). MIT Press.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(02), 109–130.

Pepperberg, I. M., & Carey, S. (2012). Grey parrot number acquisition: The inference of cardinal value from ordinal position on the numeral list. *Cognition*, *125*(2), 219–232.

Pezzulo, G., Kemere, C., & der Meer, M. A. (2017). Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Annals of the New York Academy of Sciences*, *1396*(1), 144–165.

Pfeiffer, B., & Foster, D. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 74–79.

Pfungst, O. (1907). *Das Pferd des Herrn von Osten: der kluge Hans. Ein Beitrag zur experimentellen Tier- und Menschen-Psychologie*. Leipzig: Johann Barth.

Pillsbury, W. B. (1949). James McKeen Cattell. In *Biographical Memoirs* (Vol. XXV, pp. 1–16). Washington: The National Academy of Sciences.

Pocock, R. I. (1893). Further notes and observations upon the instincts of some common English spiders. *Nature*, *49*, 60–63. https://doi.org/10.1038/049060a0

Pope, A. (1881). *Pope: Essay on Man*. Clarendon Press.

Porter, J. P. (1914). Habit formation, imitation and higher mental processes of animals. *Psychological Bulletin*, *11*(8), 294-303.

Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, *19*(1), 1–28.

Powell, N. J., & Redish, A. D. (2016). Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nature Communications*, *7*.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences*, *104*(35), 13861–13867.

Premack, D. (2010). Why humans are unique: Three theories. *Perspectives on Psychological Science*, *5*(1), 22–32.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.

Quackenbos, J. D. (1912). Is telepathy, or psychic transmission, a fact or a delusion? *The North American Review*, *196*(682), 373–382. Retrieved from https://www.jstor.org/stable/25119840

Quinsey, K. (Ed.). (2017). *Animals and Humans: Sensibility and Representation,1650-1820 2017*. Oxford: Voltaire Foundation.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107.

Ramirez-Cardenas, A., Moskaleva, M., & Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. *Current Biology*, *26*(10), 1285–1294.

Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *Journal of Neuroscience*, *24*(16), 3917–3925.

Ray, J. (1735). *The Wisdom of God Manifested in the Works of the Creation*. London: Printed by William Innys and Richard Manby. http://archive.org/details/wisdomofgodmanif00rayj

Redish, A. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, *17*(3), 147–159.

Reid, T. (1872). *The Works of Thomas Reid: Now Fully Collected, with Selections from His Unpublished Letters*. Edinburgh: Maclachlan and Stewart.

Restle, F. (1957). Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychological Review*, *64*(4), 217–228.

Rich, E. L., & Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(22), 7208–7219. https://doi.org/10.1523/JNEUROSCI.6068-08.2009

Richards, R. J. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press.

Richards, R. J. (2005). Darwin's Metaphysics of Mind." In V. Hoesle & C. Illies (Eds.). *Darwin and Philosophy*. (pp.166-180). Notre Dame University Press.

Rickard, T. C., Verfaellie, M., & Grafman, J. (2006). Transverse Patterning and Human Amnesia. *Journal of Cognitive Neuroscience*, *18*(10), 1723–1733. https://doi.org/10.1162/jocn.2006.18.10.1723

Rikhye, R. V., Gothoskar, N., Guntupalli, J. S., Dedieu, A., Lázaro-Gredilla, M., & George, D. (2019). Learning cognitive maps for vicarious evaluation. *BioRxiv*, 864421.

Rissman, J., Gazzaley, A., & D'esposito, M. (2007). Dynamic adjustments in prefrontal, hippocampal, and inferior temporal interactions with increasing visual working memory load. *Cerebral Cortex*, *18*(7), 1618–1629.

Roback, A. A. (1920). The scope and genesis of comparative psychology. *The Journal of Philosophy, Psychology and Scientific Methods*, *17*(24), 654–662.

Romanes, G. J. (1878). Animal Intelligence and Comparative Psychology. *The Spiritualist*, *13*(8), pp.1-7. Retrieved from www.iapsop.com/archive/materials/spiritualist/spiritualist_v13_n8_aug_23_1878.pdf

Romanes, G. J. (1883). *Animal intelligence*. New York, D. Appleton and Company.

Romanes, G. J. (1884). *Mental evolution in animals*. New York: D. Appleton.

Romanes, G. J. (1885). Can an animal count? *Nature*, *33*(839), 80. https://doi.org/10.1038/033080a0

Romanes, G. J. (1888). *Mental Evolution in Man: Origin of Human Faculty*. Kegan Paul, Trench.

Romanes, G. J. (1890). Cases received by the literary committee. *Journal of the Society for Psychical Research*, *4*(67), 212–213.

Romanes, G. J. (1895). *Thoughts on Religion*. Open Court Publishing Company.

Rubin, R. D., Watson, P. D., Duff, M. C., & Cohen, N. J. (2014). The role of the hippocampus in flexible cognition and social behavior. *Frontiers in Human Neuroscience*, *8*, 742. Retrieved from http://journal.frontiersin.org/article/10.3389/fnhum.2014.00742

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. PLoS Computational Biology, 13(9), e1005768.

Ryan, J. D., D'angelo, M. C., Kamino, D., Ostreicher, M., Moses, S. N., & Rosenbaum, R. S. (2016). Relational learning and transitive expression in aging and amnesia. *Hippocampus*, *26*(2), 170–184.

Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F.-X., Jbabdi, S., O'Reilly, J. X., Filippini, N., Thomas, A. G., & Rushworth, M. F. (2013). The Organization of Dorsal Frontal Cortex in Humans and Macaques. *Journal of Neuroscience*, *33*(30), 12255–12274. https://doi.org/10.1523/JNEUROSCI.5108-12.2013

Sanford, E. C. (1914). Psychic research in the animal field: Der Kluge Hans and the Elberfeld horses. *The American Journal of Psychology*, *25*(1), 1–31.

Sanz, C., Call, J., & Morgan, D. (2009). Design complexity in termite-fishing tools of chimpanzees (Pan troglodytes). *Biology Letters*, *5*(3), 293–296. https://doi.org/10.1098/rsbl.2008.0786

Sapolsky, R. M. (2004). *Why Zebras Don't Get Ulcers*. New York: Henry Holt & Co.

Scarf, D., Smith, C., & Stuart, M. (2014). A spoon full of studies helps the comparison go down: A comparative analysis of Tulving's spoon test. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00893

Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath, C. (2015). Memory and space: towards an understanding of the cognitive map. *Journal of Neuroscience*, *35*(41), 13904–13911.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., … Silver, D. (2019). Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. *ArXiv:1911.08265 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1911.08265

Schneirla, T. (1952). A consideration of some conceptual trends in comparative psychology. *Psychological Bulletin*, 49(6), 559–597.

Schneirla, T. C. (1958). The Study of Animal Behavior: Its History and Relation to the Museum. I. *Curator: The Museum Journal*, *1*(4), 17–35. https://doi.org/10.1111/j.2151-6952.1958.tb00373.x

Seamans, J. K., Lapish, C. C., & Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: Insights from electrophysiology. *Neurotoxicity Research*, *14*(2–3), 249–262. https://doi.org/10.1007/BF03033814

Secord, J. A. (2001). *Victorian Sensation: The Extraordinary Publication, Reception, and Secret Authorship of Vestiges of the Natural History of Creation* (1 edition). Chicago: University of Chicago Press.

Secord, J. A. (2015). *Visions of Science: Books and Readers at the Dawn of the Victorian Age* (1 edition). Chicago: University of Chicago Press.

Sellami, A., Abed, A. S. A., Brayda-Bruno, L., Etchamendy, N., Valério, S., Oulé, M., Pantaléon, L., Lamothe, V., Potier, M., Bernard, K., Jabourian, M., Herry, C., Mons, N., Piazza, P.-V., Eichenbaum, H., & Marighetto, A. (2017). Temporal binding function of dorsal CA1 is critical for declarative memory formation. *Proceedings of the National Academy of Sciences*, *114*(38), 10262–10267. https://doi.org/10.1073/pnas.1619657114

Selous, E. (1901). *Bird Watching*. London: J.M. Dent & Company.

Selous, E. (1905). *Bird Life Glimpses*. London: G. Allen & Company.

Shettleworth, S. (2001). Animal cognition and animal behaviour. *Animal Behaviour*, 61(2), 277–286.

Shettleworth, S. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*(11), 477–481.

Shevelow, K. (2008). *For the Love of Animals: The Rise of the Animal Protection Movement*. New York: Henry Holt and Co.

Shima, K., Isoda, M., Mushiake, H., & Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature*, *445*(7125), 315–318.

Silverman, J. L., Gastrell, P. T., Karras, M. N., Solomon, M., & Crawley, J. N. (2015). Cognitive Abilities on Transitive Inference Using a Novel Touchscreen Technology for Mice. *Cerebral Cortex (New York, NY)*, *25*(5), 1133–1142. https://doi.org/10.1093/cercor/bht293

Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children’s analogical reasoning. *Journal of Experimental Child Psychology*, *166*, 160–177.

Simon, H. A. (1956). *Administrative Behavior*. Simon and Schuster.

Singer, A. C., Carr, M. F., Karlsson, M. P., & Frank, L. M. (2013). Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. *Neuron*, *77*(6), 1163–1173.

Singer, P. (1979). Killing Humans and killing Animals. *Inquiry*, *22*(1–4), 145–156. https://doi.org/10.1080/00201747908601869

Sliwa, J., Planté, A., Duhamel, J.-R., & Wirth, S. (2016). Independent neuronal representation of facial and vocal identity in the monkey hippocampus and inferotemporal cortex. *Cerebral Cortex*, *26*(3), 950–966. https://doi.org/10.1093/cercor/bhu257

Slotten, R. A. (2004). *The Heretic in Darwin's Court: The Life of Alfred Russel Wallace*. Columbia University Press.

Small, W. S. (1899). Notes on the psychic development of the young White Rat. *The American Journal of Psychology*, *11*(1), 80–100. https://doi.org/10.2307/1412730

Smith, C. U. M., & Arnott, R. (2005). *The Genius of Erasmus Darwin*. Ashgate Publishing, Ltd.

Smith, J. D., & Church, B. A. (2018). Dissociable learning processes in comparative psychology. *Psychonomic Bulletin & Review*, *25*(5), 1565–1584.

Smith, J. D., Couchman, J. J., & Beran, M. J. (2014). Animal metacognition: a tale of two comparative psychologies. *Journal of Comparative Psychology*, *128*(2), 115–131. https://doi.org/10.1037/a0033105

Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, *79*(2), 361–374. https://doi.org/10.1016/j.neuron.2013.05.038

Sommer, A. (2012). Psychical research and the origins of American psychology: Hugo Münsterberg, William James and Eusapia Palladino. *History of the Human Sciences*, *25*(2), 23–44.

Sommer, A. (2013). Normalizing the supernormal: The formation of the "Gesellschaft für Psychologische Forschung" ("Society for Psychological Research"), c. 1886–1890. *Journal of the History of the Behavioral Sciences*, *49*(1), 18–44.

Southey, R. (1816). *The Poet's Pilgrimage to Waterloo*. Longman, Hurst, Rees, Orme and Brown, Paternoster Row.

Springer, M. S., Murphy, W. J., Eizirik, E., & O'Brien, S. J. (2003). Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proceedings of the National Academy of Sciences*, *100*(3), 1056–1061. https://doi.org/10.1073/pnas.0334222100

Stachenfeld, Kimberly L., Botvinick, M., & Gershman, S. J. (2014). Design principles of the hippocampal cognitive map. *Advances in Neural Information Processing Systems 27* (pp. 2528–2536).

Stachenfeld, Kimberly L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643-1653. https://doi.org/10.1038/nn.4650

Stanovich, K. E., West, R. F., & Toplak, M. E. (2014). Rationality, intelligence, and the defining features of type 1 and type 2 processing. *Dual-Process Theories of the Social Mind*, 80–91.

Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation. *Philosophy of Science*, *64*, S65–S74.

Sterelny, K. (2012). *The Evolved Apprentice*. MIT Press.

Stevens, J. (2017). Replicability and Reproducibility in Comparative Psychology. *Frontiers in Psychology*, *8*.

Stokes, M. G., Buschman, T. J., & Miller, E. K. (2017). Dynamic coding for flexible cognitive control. In T. Egner (Ed.) *The Wiley Handbook of Cognitive Control* (pp. 221-241). Wiley-Blackwell.

Strong, R. (2007). *Anglicanism and the British Empire, C.1700-1850*. Oxford University Press.

Stuart, T. (2007). *The Bloodless Revolution: A Cultural History of Vegetarianism from 1600 to Modern Times*. New York: W. W. Norton & Company.

Suddendorf, T. (2013a). Mental time travel: continuities and discontinuities. *Trends in Cognitive Sciences*, *17*(4), 151–152. https://doi.org/10.1016/j.tics.2013.01.011

Suddendorf, T. (2013b). *The Gap: The Science of What Separates us From Other Animals*. Basic Books.

Suddendorf, T., Bulley, A., & Miloyan, B. (2018). Prospection and natural selection. *Current Opinion in Behavioral Sciences*, *24*, 26–31.

Suddendorf, T., & Corballis, M. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*(3), 299–313.

Suddendorf, T., & Corballis, M. (2010). Behavioural evidence for mental time travel in nonhuman animals. *Behavioural Brain Research*, *215*(2), 292–298.

Suthana, N., & Fried, I. (2012). Percepts to recollections: Insights from single neuron recordings in the human brain. *Trends in Cognitive Sciences*, *16*(8), 427–436. https://doi.org/10.1016/j.tics.2012.06.006

Swisher, C. N. (1967). “The William Osler Medal Essay”: Charles Darwin on the Origins of Behavior. *Bulletin of the History of Medicine*, *41*(1), 24-43.

Sytsma, J., & Machery, E. (2012). The two sources of moral standing. *Review of Philosophy and Psychology*, *3*(3), 303–324.

Szczepanski, S. M., & Knight, R. T. (2014). Insights into human behavior from lesions to the prefrontal cortex. *Neuron*, *83*(5), 1002–1018.

Takahashi, S. (2013). Hierarchical organization of context in the hippocampal episodic code. *ELife*, *2*, e00321. https://doi.org/10.7554/eLife.00321

Takahashi, S. (2015). Episodic-like memory trace in awake replay of hippocampal place cell activity sequences. *ELife*, *4*, e08105. https://doi.org/10.7554/eLife.08105

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A map for social navigation in the human brain. *Neuron*, *87*(1), 231–243.

Taves, A. (2014). A tale of two congresses: the psychological study of psychical, occult, and religious phenomena, 1900-1909. *Journal of the History of the Behavioral Sciences*, *50*(4), 376–399. https://doi.org/10.1002/jhbs.21691

Taylor, A. H., Knaebe, B., & Gray, R. D. (2012). An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 4977–4981. https://doi.org/10.1098/rspb.2012.1998

Taylor, A. H., Miller, R., & Gray, R. D. (2013). Reply to Boogert et al.: The devil is unlikely to be in association or distraction. *Proceedings of the National Academy of Sciences*, *110*(4), E274–E274. https://doi.org/10.1073/pnas.1220564110

Thagard, P. (2010). How brains make mental models. In *Model-Based Reasoning in Science and Technology* (pp. 447–461). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15223-8_25

"The World of Psychics" (1897). *Light of Truth*, *20*, 12. Retrieved from http://www.iapsop.com/archive/materials/light_of_truth/light_of_truth_v20_n24_jun_12_1897.pdf.

Thomas, K. (1983). *Man and the Natural World: Changing Attitudes in England 1500-1800*. Allen Lane.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i–109. https://doi.org/10.1037/h0092987

Thorndike, E. L. (1899 February 25). *Letter to Cattell*. (James McKeen Cattell Papers, box 42, Thorndike, E. L.). Library of Congress, Washington D.C.

Thorndike, E. L. (1899). Do animals reason? *Popular Science Monthly*, 55, 480-490.

Thorndike, E. (1901a). The evolution of the human intellect. *Popular Science Monthly*, *60*, 58–65.

Thorndike, E. (1901b). The intelligence of monkeys. *Popular Science Monthly*, *59*, 273–279.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. Macmillan.

Tiganj, Z., Cromer, J. A., Roy, J. E., Miller, E. K., & Howard, M. W. (2018). Compressed Timeline of Recent Experience in Monkey Lateral Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *30*(7), 935–950. https://doi.org/10.1162/jocn_a_01273

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, *36*(1), 13.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. II. Place learning versus response learning. *Journal of Experimental Psychology*, *36*(3), 221.

Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189-208.

Tolman, E. (1959). Principles of purposive behavior. In S. Koch (ed). *Psychology: A Study of a Science. Vol. II.* New York: McGraw-Hill.

Tooley, M. (1972). Abortion and infanticide. *Philosophy & Public Affairs*, 37–65.

Tsutsui, K.-I., Hosokawa, T., Yamada, M., & Iijima, T. (2016). Representation of functional category in the monkey prefrontal cortex and its rule-dependent use for behavioral selection. *Journal of Neuroscience*, *36*(10), 3038–3048.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. Terrace, & J. Metcalfe (Eds.). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness* (pp. 1-57). Oxford University Press.

Turner, W. (1825). Mrs. Barbauld, *Newcastle Magazine*, 2(4): 229-232. https://search.proquest.com/openview/aaa4ba2fd02847c8/1?pq-origsite=gscholar&cbl=2589

Turner, C. H. (1909). The behavior of a snake. *Science*, *30*(773), 563–564. https://doi.org/10.1126/science.30.773.563

Turner, D. (2005). Local underdetermination in historical science. *Philosophy of Science*, *72*(1), 209–230.

Tyson, E. (1699). *Orang-Outang, Sive Homo Sylvestris: Or, The Anatomy of a Pygmie Compared with That of a Monkey, an Ape, and a Man.* London: T. Bennett and D. Brown.

Uglow, J. (2002). *The Lunar Men: Five Friends Whose Curiosity Changed the World*. New York: Farrar, Straus, and Gigoux.

Uglow, J. (2005). But what about the women? The Lunar Society's attitude to women and science, and to the education of girls. In U. Murray, & R. Smith (Ed.) *The Genius of Erasmus Darwin*. Christopher Ashgate Publishing.

Vallentin, D., & Nieder, A. (2010). Representations of visual proportions in the primate posterior parietal and prefrontal cortices. *European Journal of Neuroscience*, *32*(8), 1380–1387.

Van der Meer, M., Kurth-Nelson, Z., & Redish, A. (2012). Information processing in decision-making systems. *The Neuroscientist*, 18(4), 342–359.

Van Fraassen, B. (2008). *Scientific representation: Paradoxes of perspective*. Oxford University Press.

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes*, *78*(3), 313–334.

Voltaire, F. (1793). "Beasts." In *The Philosophical Dictionary. From the French of M. de Voltaire. A New and Correct Edition* (pp. 28-30). Dublin: Bernard Dornin.

Vom Rath, D. (1891). A contribution to the knowledge of the dermal sense-organs of the Crustacea. *Journal of Natural History*, *8*, 299-313

Von Fersen, L., Wynne, C. D., Delius, J. D., & Staddon, J. E. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*(3), 334-341.

Vonk, J., & Beran, M. J. (2012). Bears 'count' too: quantity estimation and comparison in black bears, Ursus americanus. *Animal Behaviour*, *84*(1), 231–238. https://doi.org/10.1016/j.anbehav.2012.05.001

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–956.

Wallis, E. W., & Wallis, M. H. (1903). *A Guide to Mediumship and Psychical Unfoldment*. London: Office of "Light." Retrieved from http://hdl.handle.net/2027/uc2.ark:/13960/t6057f856

Wang, H., Stradtman, G. G., Wang, X.-J., & Gao, W.-J. (2008). A specialized NMDA receptor function in layer 5 recurrent microcircuitry of the adult rat prefrontal cortex. *Proceedings of the National Academy of Sciences*, *105*(43), 16791–16796.

Warden, C. J. (1927). The historical development of comparative psychology. *Psychological Review*, *34*(2), 135–168. https://doi.org/10.1037/h0070104

Washburn, M. F. (1917). *The Animal Mind: A Text-book of Comparative Psychology*. New York: Macmillan.

Waters, R. H. (1934). The historical background of comparative psychology. In F. A. Moss (Ed.), *Comparative Psychology* (pp. 9–36). Prentice-Hall.

Watson, J. B. (1908 July 4). *Letter to Cattell*. (James McKeen Cattell Papers, box 157, Watson, John B.). Library of Congress, Washington D.C.

Watson, J. B. (1910 November 22). *Letter to Cattell*. (James McKeen Cattell Papers, box 157, Watson, John B.). Library of Congress, Washington D.C.

Watson, J. B. (1914). *Behavior: An Introduction to Comparative Psychology*. New York: H. Holt.

Watson, J. B. (1916 May 16). *Letter to Cattell*. (James McKeen Cattell Papers, box 157, Watson, John B.). Library of Congress, Washington D.C.

Watson, J. B. (1917 April 26). *Letter to Cattell*. (James McKeen Cattell Papers, box 157, Watson, John B.). Library of Congress, Washington D.C.

Watson, J. B. (1917 May 11). *Letter to Cattell*. (James McKeen Cattell Papers, box 157, Watson, John B.). Library of Congress, Washington D.C.

Watson, J. B. (1919). *Psychology from the Standpoint of a Behaviorist*. Philadelphia: Lippincott.

Weir, A. A., S, & Kacelnik, A. (2006). A New Caledonian crow (Corvus moneduloides) creatively re-designs tools by bending or unbending aluminium strips. *Animal Cognition; Heidelberg*, *9*(4), 317–334. https://doi.org/http://dx.doi.org/10.1007/s10071-006-0052-5

Wheeler, W. M. (1904). The obligations of the student of animal behavior. *The Auk*, *21*(2), 251–255.

Wheeler, W. M. (1906). The queen ant as a psychological study. *Popular Science Monthly*, *68*, 291–299.

Wikenheiser, A. M., & Redish, A. D. (2015). Hippocampal sequences and the cognitive map. In *Analysis and Modeling of Coordinated Multi-neuronal Activity* (pp. 105–129). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4939-1969-7_5

Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, *17*(8), 513–523. https://doi.org/10.1038/nrn.2016.56

Willis, T. (1683). *Two Discourses Concerning the Soul of Brutes which is that of the Vital and Sensitive of Man.* London: Dring at the Harrow.

Wilson, C. (1997). *The Invisible World: Early Modern Philosophy and the Invention of the Microscope*. Princeton University Press.

Wilson, D. A. (2002). Experimental animal behaviour studies: The loss of initiative in Britain 100 years ago. History of science, 40(3), 291-320.

Winther, R. (2006). Parts and theories in compositional biology. *Biology and Philosophy*, 21(4), 471–499.

Wirth, S., Baraduc, P., Planté, A., Pinède, S., & Duhamel, J.-R. (2017). Gaze-informed, task-situated representation of space in primate hippocampus during virtual navigation. *PLOS Biology*, *15*(2), e2001045. https://doi.org/10.1371/journal.pbio.2001045

Witmer, L. (1910). Intelligent imitation and curiosity in a monkey. *The Psychological Clinic*, *3*(8), 224–227.

Wolffram, H. (2006). Parapsychology on the couch: the psychology of occult belief in Germany, c. 1870–1939. *Journal of the History of the Behavioral Sciences*, *42*(3), 237–260.

Wolffram, H. (2009). *The Stepchildren of Science: Psychical Research and Parapsychology in Germany, c. 1870-1939*. Brill.

Wu, C.-T., Haggerty, D., Kemere, C., & Ji, D. (2017). Hippocampal awake replay in fear memory retrieval. *Nature Neuroscience*, *20*(4), 571–580. https://doi.org/10.1038/nn.4507

Wundt, W. M. (1896). *Lectures on Human and Animal Psychology*. New York: Macmillan & Co.

Wynne, C., & Udell, M. (2013). *Animal Cognition: Evolution, Behavior and Cognition*. Palgrave Macmillan.

Yerkes, R. M. (1901). The formation of habits in the turtle. *Popular Science Monthly 58*, 519-525.

Yerkes, R. M. (1914). The Harvard laboratory of animal psychology and the Franklin field station. *Journal of Animal Behavior*, *4*(3), 176–184. https://doi.org/10.1037/h0076089

Zheng, T., & Wilson, C. (2002). Corticostriatal combinatorics: the implications of corticostriatal axonal arborizations. *Journal of Neurophysiology*,87(2), 1007–1017.