In Defense of Instinct Concepts

Author: Kelle Dhein

Affiliation: Arizona State University

Book Title: *Life and Mind - Theoretical and Applied Issues in Contemporary Philosophy of Biology and Cognitive Sciences*

Chapter Title: "In Defense of Instinct Concepts"

Keywords: Instinct versus Learning, Nature versus Nurture, Nativism versus Empiricism, Innate versus Acquired, Animal Behavior

Abstract: In the 20th century, the distinction between instinct and learning motivated international debates that reshaped the disciplinary landscape of animal behavior studies. When the dust settled, a new consensus emerged: the development of behavioral traits involves complex interactions between organism, genetic inheritance, and experience with the environment. This insight has spurred some philosophers and scientists to eschew instinct versus learning dichotomies—and instinct concepts in particular—on epistemic grounds. In this paper, I reassess influential 20th century arguments against instinct concepts and instinct vs. learning dichotomies to show that these arguments have limited scope. Then, I use historical case studies to demonstrate the combinatorial flexibility of instinct and learning concepts. Although instinct and learning are often framed as mutually exclusive opposites, scientists continue to combine them in causal physiological accounts of behavior. I conclude by suggesting that instinct concepts help scientists achieve their epistemic aims because of the way they facilitate abductive inferences.

Keywords: Instinct versus Learning, Nature versus Nurture, Nativism versus Empiricism, Innate versus Acquired, Animal Behavior

1. Introduction

In the 20th century, the distinction between instinct and learning motivated international debates that reshaped the disciplinary landscape of animal behavior research (Hinde 1966; Aronson et al. 1970; Beer 1975, 2020; Boakes 1984; Barlow 1991; Dewsbury 1992, 2002; Griffiths 2004; Marler 2004; Brigandt 2005; Burkhardt 2005; Feest 2005; Burghardt 2009; Dhein 2021). When the dust settled, a new consensus emerged: the development of behavioral traits involves complex interactions between organism, genetic inheritance, and experience with the environment. This interactionist insight has spurred some philosophers and scientists to eschew instinct versus learning dichotomies—and instinct concepts in particular—on epistemic grounds (Hebb 1953; Lehrman 1953, 1970; Beach 1955; Oyama 2000; Griffiths 2002; Robinson 2004; Mameli and Bateson 2006; Griffiths et al. 2009; Shea 2012; Linquist 2018; Vicedo 2023). In this paper, I defend instinct concepts and instinct versus learning dichotomies against instinct eliminativists.

But before getting into my defense, it is worth reflecting on why the epistemic value of instinct concepts matters at all. The most obvious answer is that it matters for good science. Adopting Midgley's (1992) view of philosophy as conceptual plumbing, instinct concepts are like an old and unfamiliar system of pipes and fittings hidden beneath the floorboards of the behavioral sciences. Many suspect that improvements in plumbing over the last century have made these pipes and fittings obsolete. I argue that water still flows through them, and I suggest that the behavioral sciences are better for it. So, the most straightforward motivation for this paper is to contribute to the general upkeep of the conceptual plumbing supporting the behavioral sciences by combating a common misconception. Namely, that instinct concepts are faulty plumbing distorting scientists' thinking about behavior.

More broadly, as historians, sociologists, and philosophers of science are at pains to show, science is not separate from society. When scientists develop concepts, they are influenced by the conceptual plumbing of the society in which they live. In turn, concepts developed in scientific contexts seep through and influence larger society. This is especially true for the behavioral sciences because so many important everyday situations force people to confront problems involving behavior (e.g. "Why won't my boss give me a raise?"). The behavioral sciences also speak to peoples' desire to understand themselves. If one had to translate the philosophical maxim "Know thyself" into a scientifically tractable problem, one could do much worse than "Why do I behave the way I do?" Scientific concepts that promise to help explain, predict, or control behavior concern everyone because they are bound to reenter society and color the way we deal with ourselves and others.

I use the phrase "instinct versus learning" throughout this paper, but I could have used equivalent slogans like "nature versus nurture", "nativism versus empiricism", "inheritance versus environment", or "innate versus acquired". These phrases carry different connotations and enjoy varying levels of popularity across research communities over time. Still, they refer to a common distinction that has been a persistent theme throughout the history of biology. This paper is concerned with the role of this conceptual distinction (and instinct in particular) in the behavioral sciences, not the words scientists use to invoke that distinction. Some philosophers have argued that scientists should do away with the terms "instinct" or "innate" because those terms have folk meanings that may bleed into the science (Griffiths 2002; Knobe and Samuels 2013; Linquist 2019). A background motivation for this paper is a more forward-looking worry about scientific meanings bleeding back into society. If instinct concepts have unrecognized epistemic value, as I contend, then a resurgence of popular instinct-thinking is plausible. For most, this prospect is not comforting, especially given the historical connections between instinct, biological determinism, and racism. Philosophers and historians of science have a special role to play here because they clarify the meanings of scientific concepts over time and space. A clear understanding of how scientific concepts work in scientific contexts is necessary for achieving appropriate societal interpretations of those concepts.

My defense of instinct versus learning dichotomies and instinct concepts in particular takes three steps. First, I reassess two 20th century arguments against instinct to show that these arguments have limited scope. To critique the first argument, I show that it only applies to research aimed at the developmental dimension of behavior, not research on the causal physiological dimension. To critique the second argument, I show that it only applies to certain patterns of scientific inference, not all patterns.

My second step is to adopt a use-based view of scientific concepts wherein the content of a concept is tied to the ways scientists use that concept to achieve their epistemic aims (Brigandt 2010, 2012; Feest 2010, 2012; Arabatzis 2012; Boon 2012; Kindi 2012; Steinle 2012; Margolis and Laurence 2013; Smith 2020). A key upshot of this view is that the instinct versus learning dichotomy is not a monolith. Because different research groups have used the concepts in different ways throughout the history of the behavioral sciences, there are a plurality of instinct and learning concepts. Recognizing this diversity helps reveal the surprising ways scientists continue to combine instinct and learning to create causal physiological accounts of behavior. In short, I weaken arguments against the instinct side of instinct vs. learning dichotomies and then demonstrate the dichotomies' continued relevance by examining different ways scientists from different traditions combine instinct and learning.

My third step is to suggest a positive epistemic role for instinct concepts in the behavioral sciences: instinct concepts facilitate abductive inferences in a way that guides and constrains hypothesis formulation. The idea is that "instinct" is an umbrella concept containing different behavioral properties. When scientists discover that a behavior of interest exhibits some property under the instinct umbrella, they can engage in instinct-thinking by abductively inferring that the behavior also possesses other properties under the instinct umbrella. These inferences become hypotheses when scientists design experiments to test them. Instinct-thinking has epistemic value insofar as it guides the investigative trajectories of scientific communities towards useful research outputs.

After surveying philosophical work on the innate/acquired distinction, Griffiths and Linquest (2022, p. 26) write, "Although one cannot deny that this concept [innateness] has been productive, it is a further question whether certain sciences would have advanced more rapidly if they had not framed their hypotheses in terms of the innate/acquired distinction." This is an important question for my line of argumentation. Future work in counterfactual histories of science may be able to address it (Jamieson and Radick 2013; Radick 2005, 2016). Griffiths and Linquist (2022) also conclude that "an adequate philosophical diagnosis of what is at issue in scientific debates over nativism should sample from a broad range of historical periods and figures." This paper is an initial attempt to realize that goal.

The rest of the paper proceeds as follows. In section 2, I argue that 20th century critiques of instinct have a smaller scope than they claim. In section 3, I demonstrate the combinatorial flexibility of instinct vs. learning dichotomies with multiple case studies. In section 4, I suggest instinct concepts continue to help behavioral scientists achieve their epistemic aims.

2. 20th Century Arguments against Instinct and Instinct versus Learning Dichotomies

Instinct concepts suffered multiple setbacks within the behavioral sciences in the 20th century. In cultural, political, and social domains, they became associated with eugenics, Nazi science, and controversial attempts to explain human social phenomena in biological terms (Lewontin 1979; Barlow 1991; Senchuk 1991; Griffiths 2002, 2004, 2008). Epistemologically, some suspected that instinct concepts retained unscientific elements from their past development in 19th and 18th century theories of behavior (Dunlap 1919; Bernard 1921; Kuo 1921; Burghardt 1973, p. 323; Stich 1975; Griffiths 2002). However, the most significant 20th century arguments against the epistemic value of instinct in the behavioral sciences stem from American experimental psychologists.

Daniel Lehrman (1953) produced the most influential articulation of these critiques, which are widely regarded as a distillation of his doctoral mentor, Theodore Schneirla's, approach to studying behavior. Tracing the lineage back another generation, historian of science Donald Dewsbury (2002) includes Schneirla with the Chicago Five, a group of integrative psychobiologists who worked with Karl Lashley between 1929–1935 and whose research reflects a common set of underlying themes. One underlying theme common to the Chicago Five that works its way into Lehrman's critique of instinct is that "Behavioral development is an epigenetic process resulting from the continuous, dynamic interaction of genes, environment, and organisms" (Dewsbury 2002, p. 28). Lehrman's 1953 paper directs this line of thinking against Austrian ethologist Konrad Lorenz's theory of instincts.[1]

Lehrman's critique of Lorenzian instinct has received much attention in the historical, philosophical, and scientific literature (Beer 1975, 2020; Dewsbury 1992; Rosenblatt 1995; Griffiths 2002; Marler 2004; Burkhardt 2005; Griffiths and Linquist 2022; Vicedo 2023). Rather than rehearse all Lehrman's arguments, I focus on two especially influential arguments that have been applied to multiple instinct concepts, not just Lorenz's. Then, I show that these arguments have narrower scopes than instinct eliminativists maintain.

First, consider Lehrman's (1953) argument that Lorenz's instinct concept is epistemically detrimental to behavior science because it obfuscates the need for detailed accounts of how inheritance and environment interact with a developing organism to produce behavioral traits. This argument is the foundation of interactionist critiques of instinct. There are at least three ways Lehrman argues the obfuscation of development leads to bad epistemic outcomes. The first is that instinct acts as a pseudo explanation that deflects inquiry away from developmental processes. When scientists label a behavior as instinctive, it causes them to ignore questions of development and treat the behavior as given. Second, Lorenz's instinct concept posits a hard line dividing inherited instinctive behavior from behavior learned through experience. Lehrman argues that Lorenz's hard distinction implies that some behavioral traits develop without environmental input. However, as Lehrman and the Schneirla school repeatedly emphasized, the

[1] For historical accounts of the development of Lorenz's instinct theory, see Schleidt (1962) and Brigandt (2005).

nuanced intricacies of developmental processes and mechanisms make that implication highly unlikely.

So far, Lehrman's arguments have been against the instinct versus learning dichotomy in the sense that a dichotomy requires two paired elements, and Lehrman is attacking one of those elements (i.e. instinct). The third way Lehrman argues that instinct obfuscates behavioral development is directed at the whole dichotomy. Rather than conceive of development as the interaction of inheritance and environment, Lehrman argues that scientists should replace that dichotomy with the interaction between organism, the organism's internal environment, and the organism's external environment. An epistemically favorable upshot of Lehrman's reframing is supposed to be a renewed emphasis on the processual nature of development.[2]

These critiques of instinct and the instinct vs. learning dichotomy only apply to research aimed at the developmental dimension of behavior. Put differently, all the bad epistemic outcomes enumerated above are bad because they hinder scientists from solving problems about behavioral development. I need to make this seemingly obvious point because the gist of Lehrman's interactionist critique is often repeated but rarely qualified. Famously, the Dutch ethologist Nikolaas Tinbergen (1963) proposed four dimensions of behavior that are amenable to scientific inquiry: Development, Causation, Survival Value, and Phylogeny. Although the interactionist critique is often taken to be an argument against instinct in general, it is limited to

---

[2] "The problem of development is the problem of the development of new structures and activity patterns from the resolution of the interaction of *existing* structures and patterns, within the organism and its internal environment, and between the organism and its outer environment. At any stage of development, the new features emerge from the interactions within the *current* stage and between the *current* stage and the environment. The interaction out of which the organism develops is *not* one, as is so often said, between heredity and environment. It is between *organism* and environment! And the organism is different at each different stage of its development" (Lehrman 1953, p. 345, italics in original).

behavioral research with the epistemic aims of development. To avoid confusion, I will refer to Tinbergen's category of "causation" as "causal physiological" since cause and effect relationships are relevant for investigating any of the dimensions of behavior. In section 4, I suggest that instinct concepts have epistemic value for solving problems about the causal physiological dimension of behavior.

Before moving on, it's worth noting that Lehrman's (1953) writing is sometimes ambiguous so that one can interpret him as conflating questions of development and more proximate physiological causation. Lehrman (1953, p. 345, italics in original) writes,

> "The problem for the investigator who wishes to make a causal analysis of behavior is: How did this behavior come about? The use of 'explanatory' categories such as 'innate' and 'genically [sic] fixed' obscures the necessity of investigating developmental *processes* in order to gain insight into the actual mechanisms of behavior and their interrelations."

Here, Lehrman equates the "causal analysis of behavior" with the question, "How did this behavior come about?" But the phrase "how behavior comes about" has a developmental and causal physiological interpretation. He also claims that investigating developmental processes is necessary to uncover "the actual mechanisms of behavior," and it's unclear whether "actual mechanisms" are developmental mechanisms, more proximate physiological mechanisms, or both.

On the most charitable reading, Lehrman's use of the phrases "causal analysis of behavior", "how behavior comes about", and "the actual mechanisms of behavior" refer to analyzing the causal dynamics of *developmental* mechanisms, not causal physiological mechanisms. The distinction matters because research programs aimed at developmental mechanisms tend to involve different investigative practices, theoretical commitments, experimental subjects, and epistemic aims than research programs aimed at causal-behavioral mechanisms. For instance, a research group may focus on the neurophysiological mechanisms underlying adult toad tongue-shooting behavior (Ewert 1970, 1987; Ewert et al. 1983, 1999). This is a different epistemic aim that entails different investigative practices than that of illuminating the developmental mechanisms that were preconditions for the existence of the neurophysiological mechanisms responsible for adult toad tongue-shooting. Of course, as Tinbergen (1963) himself argued, all the dimensions of behavior are interrelated. Thus, insights into the ontogeny of a behavioral trait can plausibly help scientists investigate more proximate causal physiological mechanisms underpinning the trait's performance (and vice versa). Still, these two kinds of research are aimed at distinct goals, and Lehrman's interactionist argument only applies to research aimed at problems of development.

Lehrman's second influential critique of instinct maintains that the concept licenses invalid inferences. The idea is that the concept of instinct contains various instinctive properties. When scientists identify a behavioral trait as instinctive because the trait possesses one instinctive property, they erroneously infer that the trait also possesses other instinctive properties. For example, Lorenz's theory of instinct held that instinctive behaviors are developmentally fixed, meaning that developmental processes robustly output the behavioral trait regardless of environmental variation. Lorenz also (1937; 1950) proposed a psycho-hydraulic model of instinct to represent the functional dynamics of neurophysiological mechanisms underlying instinctive behavior. Lehrman's point is that such properties need not co-occur. If a scientist discovers a developmentally fixed behavior, they are not justified in inferring that the behavior is also produced according to dynamics represented by Lorenz's model. Historically, this argument against instinct has focused on the properties of being widespread throughout a species, of having been evolved to realize an adaptive outcome, and of being developmentally fixed (Griffiths 2002). More recently, critics of instinct have proposed longer lists of properties that define instinct concepts (Mameli and Bateson 2011; Shea 2012; Linquist 2019). The argumentative upshot within these papers is the same: instinct concepts are epistemically deleterious because they facilitate invalid inferences between instinctive properties.

According to this critique, inductively judging a behavior to be instinctive erroneously facilitates further deductive inferences about the behavior. As with the first argument against instinct, I need to emphasize the scope of this critique. There are many more patterns of scientific inference potentially relevant to instinct concepts than the one just described. In section 4, I suggest that instinct concepts facilitate a different pattern of scientific inference, and that these inferences help scientists achieve their epistemic goals.

Brigandt's (2010) three-part analysis of scientific concepts helps summarize this section. Brigandt (2010) argues that the contents of scientific concepts can be characterized in terms of (1) what the concept refers to, (2) what inferences the concept facilitates, and (3) the epistemic goal pursued with the concept's use. In this section, I have argued that instinct concept eliminativists construe (2) and (3) too narrowly.

3. Case Studies Demonstrating the Combinatorial Flexibility and Continued Relevance of Instinct Concepts

In this section, I use case studies to demonstrate the combinatorial flexibility and continued relevance of instinct concepts. Although instinct and learning are often framed as mutually exclusive opposites, these case study show how the two concepts complement each other within causal physiological accounts of behavior. Broadly put, the problem of producing causal physiological accounts of behavior is the problem of explaining how a behavioral system puts the right parts in the right places at the right times. Scientists generally assume that these behavioral events are directed at goals and that they are initiated by discoverable causes. Below, I show how different scientists posit instinctive properties at different causal junctures of behavioral events. I adopt a use-based view of scientific concepts wherein the content of a concept is tied to the way scientists use the concept (Peirce 1878; Wittgenstein 1953; Sellars 1968; Brandom 1994; Brigandt 2010). This use-based method of demarcating concepts matters because it reveals the diversity of instinct concepts over time and space. Instinct concepts play various roles in causal physiological accounts of behavior.

3.1. Historical Case Study: Two Developmental Paths for Wallace Craig's Theory of Instincts

Dhein (2023) traces the cognitive map debate in insects back to the publication of American animal psychologist Wallace Craig's (1918) paper, "Appetites and Aversions as Constituents of Instinct." Here, Craig advances an influential theory of instincts that was adopted and modified in different ways by American animal psychologist Edward Tolman (1932) and Austrian ethologist, Konrad Lorenz (1937).

Craig's (1918) theory of instincts holds that animals enter into appetitive states and aversive states. Appetitive states agitate the animal until a specific stimulus is received and aversive states agitate the animal until a specific stimulus is removed. Appetitive and aversive behaviors are purposeful in the objective sense that they tend to continue until the relevant stimulus is received or removed. Appetitive and aversive behaviors can be shaped by learning so that they become more effective at receiving/removing relevant stimuli. Once the relevant stimulus is received or removed, it triggers what Craig calls a "consummatory reaction". The consummatory reaction is a rigid, stereotyped, and reflex-like chain of action. After a consummatory reaction is triggered, the animal returns to a state of rest.

In a recent reevaluation of Craig's (1918) paper, Burghardt and Burkhardt (2018) highlight two important features of Craig's instinct theory. First, they note that the theory is primarily related to "matters of behavioral causation" (Burghardt and Burkhardt 2018, p. 362). This is the same distinction I made in section 2 when I cited Tinbergen's (1963) four questions paper to contrast the developmental and causal physiological dimensions of behavior. Craig's instinct theory offers a functional model of the more proximate psycho-physiological dynamics underlying behavioral performances. It does not aim to illuminate the developmental processes that made the animal the sort of system that realizes Craig's functional model.

The second important feature of Craig's theory that Burghardt and Burkhardt (2018) highlight is the way it distinguishes two different conceptions of instinct while incorporating both. The first appetitive/aversive phase of Craig's theory uses motivational, energetic, or drive-based conceptions of instinct. The second phase of Craig's theory, the triggering of a consummatory reaction, uses reflex-based conceptions of instincts as stereotypical chains of action. These two conceptions of instinct are characterized by different behavioral properties. What these different behavioral properties share in common is that they are all causal physiological properties, not properties characterizing the developmental dimension of behavior.

Craig's (1918) motivational conception of instinctive behavior includes the property of being goal-directed in the objective sense described above. It also includes the property of being caused by internal neurophysiological processes whose functional dynamics can be usefully described in terms of "energy" flowing through "channels", "accumulating", and "discharging". Alternatively, Craig's reflex-based conception of instinctive behavior includes the property of being highly stereotyped. And it includes the property that the animal need not be aware of the adaptive outcome caused by the consummatory reaction. For an animal, the performance of the consummatory reaction is a goal in itself. The performance is sufficient to enter a state of rest.

Although Craig's (1918) theory is usually remembered as an instinctive theory of behavior, Craig's motivational conception of instinct is also compatible with learning. Appetitive and aversive motivational states are like an engine that drive and direct behavior towards or away from stimuli. Though motivated by instinct, such behaviors can be shaped by learning over time. Craig's motivational conception of instinct provides necessary success/failure criteria for trial-and-error learning. The important points are that Craig's theory is aimed at elucidating the causal physiological dimension of behavior and that Craig's theory features a configuration of multiple instinct concepts and learning.[3]

Scientists and historians have argued that Craig's (1918) theory of instincts lives on in contemporary behavior systems theory (Burghardt and Bowers 2017; Burghardt and Burkhardt 2018; Burghardt 2020). Below, I focus on the way Craig's (1918) theory influenced Tolman and Lorenz. As shown by Dhein (2023), Tolman (1932) and Lorenz (1937) adopted different aspects of Craig's (1918) theory for different purposes.

Tolman (1932) transforms Craig's motivational notion of instinct into "ultimate drives" which are ultimate in the sense that they motivate all behavior (Feest 2005). Tolman uses the

[3] For an older example of an instinct/learning configuration, consider Patton's (2022) recent examination of the Helmholtz/Hering debate, a debate often framed in terms of empiricism (Helmholtz) versus nativism (Hering). As shown by Patton, Hering's nativist position regarding visual perception incorporated learning via the Lamarckian notion of "organic memory".

goal-directed notion of purpose provided by Craig's motivational instinct concept to bring the teleological notion of purposiveness into experimental learning studies. As Dhein (2023), summarizes, for Tolman, *instincts drive learning*. The rat-in-maze learning experiments that were so characteristic of Tolman's research are designed to produce evidence about the more variable appetitive phase of behavior (i.e. learning to navigate a maze to receive/remove goal stimuli), not the rewarding consummatory reactions rats perform when they receive/remove such stimuli in naturally-occurring contexts (e.g. consuming food) (Burghardt and Burkhardt 2018, p. 367). It would require deeper analysis to confirm, but the motivational notion of instinct in the background of Tolman's theory may have influenced his investigative practices. Rat-in-maze experiments have become emblematic of 20th century American learning-heavy approaches to animal behavior, but as Tolman's case suggests, they may bear subtle marks of instinct-thinking. Furthermore, this Tolmanian research tradition has been quite successful. Tolman's (1948) paper, "Cognitive Maps in Rats and Men", first proposed the cognitive map hypothesis based on the results of rat-in-maze experiments. In 2014, the Nobel Prize in Physiology or Medicine went to scientists who had developed that hypothesis and research tradition to provide a causal, neuro-physiological account of how cognitive maps work in rats (Burgess 2014).

Lorenz (1937) largely focused on the reflex-based notion of instinct embodied in Craig's (1918) consummatory reaction. For Lorenz, the appetitive/aversive phase of behavior is not an instinct, only the consummatory reaction is instinctive. As one would expect, Lorenz characterized his conception of instinct with some of Craig's reflex-based properties. For instance, Lorenz's instinctive behaviors consist of invariant, stereotyped movements. Following Craig, Lorenzian instinct holds that animals are not aware of the adaptive outcomes their instinctive behaviors are aimed at. As discussed in section 2, Lorenz's instinct concept also included developmental properties that were the target of Lehrman's first critique. More surprisingly, Lorenz includes a property from motivational conceptions of instinct. Namely, Lorenz argues that the causal mechanisms underlying instinctive behaviors share a common functional structure that can be usefully understood in terms of energy flowing through channels, accumulating, and releasing (See Figure 1).

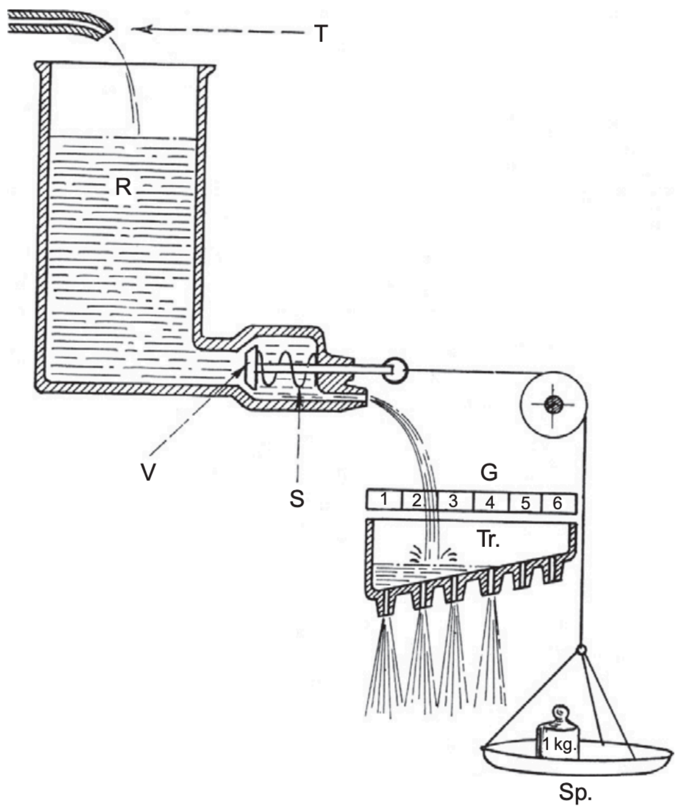

Figure 1: Illustration of Lorenz's (1950) psycho-hydraulic model of instinctive behavior. Although this illustration was published over a decade after Lorenz's (1937) instinct paper, it remains a useful tool for visualizing the way Lorenz (1937) conceives of instincts. The tap T supplies a constant flow of liquid representing the endogenous production of action-specific energy. Reservoir R represents the amount of this energy that has built up in the animal. Cone valve V represents the instinct-releasing mechanism, and spring S represents inhibitory pressure stopping the instinct from being released. Pan SP represents the perceptual aspects of the instinct-releasing mechanism, and the 1 kg weight represents impinging stimulation (the heavier the weight, the more intense the releasing stimuli). The instinctive behavior pattern is the jet of liquid pouring out of the reservoir, and measuring stick G indicates the intensity of the behavior pattern. The slanted bottom of trough Tr shows how different intensities of instinctive behavior patterns cause sequences of different activities. Description from Dhein (2023).

Lorenz stands out among instinct theorists for the hard line he draws between learning and instinct. Still, Lorenz's division of learning and instinct into totally discrete components of behavior does not preclude the possibility that these components interact (Eibl-Eibesfeldt 1961). Lorenz argued that instinct concepts could account for sophisticated behaviors, but he did not argue that all behavior is instinctive. Like Craig's appetitive and aversive states, Lorenz's accounts of behavior have room for learning. As Dhein (2023) summarizes, for Lorenz, animals learn to trigger instincts.

A key difference between the way Tolman and Lorenz use instinct and learning concepts is the way they configure instinct and learning concepts in their accounts of behavior. Thinking about the experiments they performed highlights this point. For Tolman, the interesting part of behavior is the way it can be shaped by experience. His rats are interesting when they use past experience of maze running to guide current maze running. The rats are not as interesting once they reach the end of the maze and encounter their reward.[4] This way of thinking has seeped into broader society to such an extent that it is now difficult to adopt the Lorenzian perspective, which refutes the assumption that learned behaviors are more sophisticated or interesting than instinctive behaviors. Remember that the problem of producing causal physiological accounts of behavior is the problem of explaining how a behavioral system puts the right parts in the right places at the right times. The Lorenzian perspective holds that reflex-based notions of instinct can do more of that explanatory work than learning theorists allow. One of the ways reflex-like

[4] Consider Tolman's (1938, p. 34) conclusion to a review paper on the significance of the rat-in-maze experimental paradigm: "Let me close, now, with a final confession of faith. I believe that everything important in psychology (except perhaps such matters as the building up of a super-ego, that is everything save such matters as involve society and words) can be investigated in essence through the continued experimental and theoretical analysis of the determiners of rat behavior at a choice point in a maze."

instincts accomplish this work, according to Lorenz, is by offloading the task of behavioral coordination to the environment. As Griffith (2008, p. 395–96) remarks,

> "A striking feature of Lorenz's instinct theory is that the coordination of instinctive behavior into effective sequences is dependent on the distribution of releasing stimuli in the organism's natural environment. Although each specific instinct – collecting twigs at nesting time, inserting twigs into the nest and so forth – corresponds to a neural mechanism, the larger structure of instinctual behavior only emerges in the interaction between those mechanisms and the organism's natural environment. The environment has thus taken over the role of nebulous coordinating forces like the 'nesting instinct' postulated by earlier instinct theories."

As a result, Lorenz's well-known experiments with Tinbergen emphasize putting animals in environments that retain relevant stimuli from their naturally occurring environments.[5]

Although Tolman is broadly associated with learning and Lorenz with instinct, their theories of behavior each retain an element from the other side of the dichotomy. However, there is a sense in which Lorenz's inclusion of learning and the environment and Tolman's inclusion of instinct are trivial and unremarkable. For something to behave, it must possess preexisting structures with which to behave. To understand behavior, scientists must consider the context in which it is performed. These examples of how Lorenz and Tolman drew from the same instinct theory to build such divergent theories of behavior are not just meant to exhibit the compatibility of instinct and learning in scientific accounts of behavior. They are meant to exhibit the

---

[5] In Hirsch et al. (1955), Tolman and colleagues critique Tinbergen and Lorenz's hawk/goose cutout experiments. Hirsch et al.'s (1955) argumentative strategy is representative of 20th century nature/nurture debates in that they attempt to refute the innateness of a behavioral trait by demonstrating that the trait requires experiential input to develop properly.

combinatorial flexibility of learning and instinct concepts. Philosophers and scientists seem to underappreciate how large the possibility space of instinct/learning combinations is for problems of physiological causation.

In Linquist's (2018, p. 12) critique of innateness, he asks whether innateness is associated with a productive research program and how cognitive science research programs that are more wedded to innateness concepts compare to those that are not. Superficially, the Tolman/Lorenz split in the 1930s looks like a split between productive science and unproductive science with the productive side being learning-heavy. As mentioned earlier, Tolman's research tradition led to productive work on cognitive maps. Lorenz, on the other hand, has a much more contested legacy. Historically, Lorenzian instinct is often framed as a flawed aspect of classical ethology that had to be abandoned for the eventual synthesis of ethology and comparative psychology during the last half of the 20th century.[6] Lehrman's arguments against Lorenzian instinct have played a large part in justifying that historical framing. In the next subsection, I complicate this narrative by showing how contemporary behavioral scientists continue to propose causal physiological accounts of behavior featuring Lorenzian hydraulics.

3.2. Learning with Lorenzian Hydraulics

Given Lorenz's reputation as a champion of instinct, a neo-Lorenzian theory of learning may seem incoherent. However, at least two separate communities of contemporary behavioral scientists have proposed models of behavior that combine aspects of Lorenzian instinct and 20th century learning theory. The first group uses behavioral experiments to investigate the

[6] For rare perspectives highlighting the value of Lorenzian instinct, see Eibl-Eibesfeldt (1961; 1997), Marler (2004), and Ronacher (2019).

neurophysiological mechanisms underpinning insect navigation and locomotion. The second group builds dynamical models of human decision making to inform health policy interventions. I begin with insect navigation.

One common strategy that insects like the desert ant *Cataglyphis* use to navigate is called path integration. Path integration allows navigators to return to a point of origin via the most direct route no matter how haphazardly they wander. To navigate via path integration, a navigator needs some way of recording information about the distance and direction they travelled from the point of origin (e.g. an ant nest). Formally, modelers represent this information about an animal's outgoing journey as vectors. Integrating all the vectors representing an animal's outgoing journey produces a single vector that represents the most direct path from the animal's current location to the origin point. Ants that do path integration use their experience of walking away from their nest to guide the way they walk back to their nest.

Cataglyphis foragers have multiple ways of recording how far they have walked. They can monitor optic flow, or the rate at which environmental features move across their field of vision. They can also rely on the regularity of their gait to do something like step-counting or stride integration. Presumably, these methods for recording distance travelled operate in tandem. Wolf et al. (2018) designed cue conflict experiments to investigate the functional organization of the neurophysiological mechanisms responsible for recording distance travelled in *Cataglyphis* foragers.

In Wolf et al.'s (2018) cue conflict experiments, they manipulated the cues associated with each distance-recording mechanism so that the optic flow and stride integrator output different values for distance traveled. As a result, manipulated ants often took comprise routes on their inbound journeys somewhere between the two conflicting routes indicated by their distance estimating mechanisms. By comparing the routes ants travel when the values of their distance estimating-mechanisms are mismatched in different ways, Wolf et al. (2018) produce evidence about how these mechanisms interact to guide navigation. To make sense of that evidence, Wolf et al. (2018) adopt a Lorenzian concept of instinct (See Figure 2).

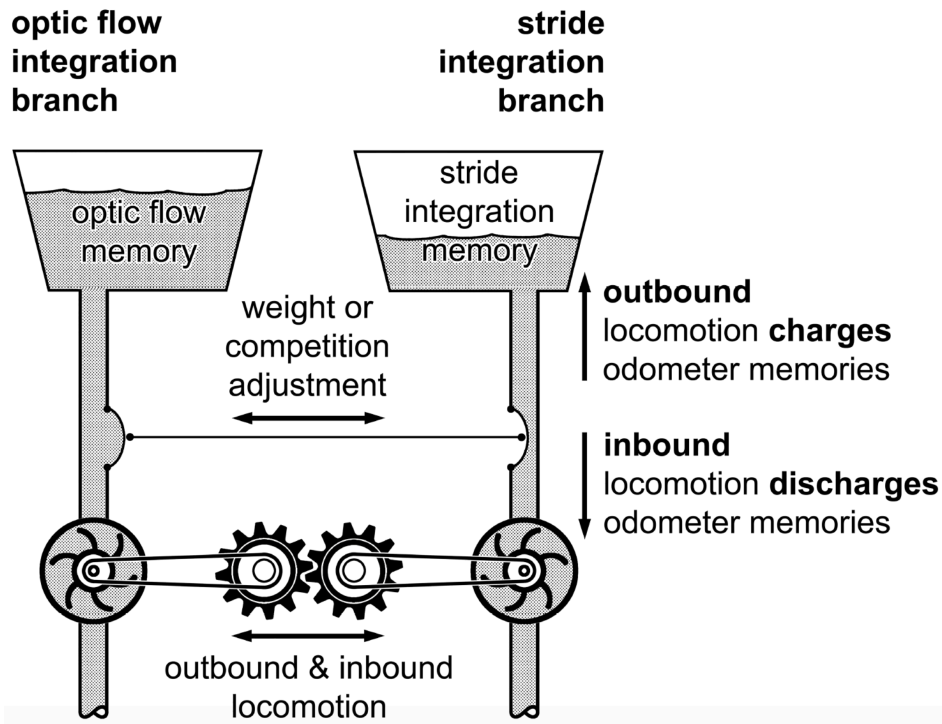

Figure 2: Wolf et al.'s (2018) hydraulic model of optic flow integration and stride integration odometry in *Cataglyphis* ants.
"Outbound and inbound locomotion (symbolised by cogwheels, bottom centre) drive the optic flow and stride integrators (turbine pumps, driven by cogwheels). The distance integrators fill or deplete their respective distance memories (elevated water troughs above turbine pumps) during outbound or inbound travel, respectively (arrows on the right). Either odometer may work alone when the other one is incapacitated (cogwheels would be disengaged, only one of them active). Normally, however, the two odometer memories drive homing together, although with different weights, or competitively. This is illustrated by two interconnected membrane valves that open or close the supply pipes to the water troughs at each other's expense. Competitive interaction is indicated by the valve to the trough with the higher water level, and thus higher water pressure, constricting the valve in the pipe with the lower pressure. Although both odometers would normally measure the same walking distance, the (symbolic) water levels need not to be identical. Neither the denomination nor even the currency of the two odometer memories have to be identical" (Wolf et al. 2018, p. 11)

Wolf et al. (2018) explicitly frame their hydraulic model as a descendant of Lorenz's (1937, 1950). The most striking continuity is the analogy between fluid dynamics and behavior. However, the liquid in Lorenz's model represents action-specific energy while Wolf et al.'s (2018) liquid represents memory. This is a significant reshuffling of instinct and learning concepts. Lorenz's theory of instincts was compatible with learning. Like Craig, he allowed that animals could learn to trigger instincts. But in Wolf et al. (2018), instinct and learning are more deeply intertwined. Broadly put, Wolf et al.'s (2018) model posits that *insects use the products of learning in an instinct-like way*. More specially, Wolf et al. (2018) use a motivational notion of instinct to characterize the functional, neurophysiological dynamics that allow memories to guide action. This is also significantly different than Tolman, who used a motivational notion of instinct to bring purposive directedness to learning. 20th century learning theorists like Tolman tended to describe the way memories influence action in terms of "associations" and "inferences". Wolf et al. (2018) use instinct to describe that relationship in terms of memories being "charged" and "depleted" or "discharged" in ways that trigger different behaviors. The general point here is that Wolf et al. (2018) provides a recent demonstration of the relevance and combinatorial flexibility of instinct and learning concepts for causal physiological accounts of behavior.

The final example also involves a hydraulic model of behavior. In 2014, a group of control systems engineers and health researchers created a hydraulic dynamical systems model of human health behavior according the principles of social cognitive theory (Martín et al. 2014) (See Figure 3). The model is dynamical in the sense that it takes inputs, gives outputs, observes consequences of outputs, and then adapts according to those consequences over time. Whereas Wolf et al.'s (2018) model encapsulates a hypothesis that suggests further behavioral experiments in what many would consider a “basic science” research program, Martín et al.’s (2014) model simulates the behavior of human agents in different situations to inform health policy interventions. Martín et al.’s (2014) model is more removed from behavioral experiments in the sense that their model is informed by a longstanding, much elaborated theory of behavior that is in turn informed by behavioral experiments.[7] The sorts of behavioral experiments that informed social learning theory and then social cognitive theory came from the learning side of the 20th century instinct/learning divide (e.g. Skinner 1953; Bandura 1986). Like Wolf et al.’s (2018) model, Martín (2018) use a motivational notion of instinct to characterize the functional dynamics of learning mechanisms.

[7] One important exception is the special attention Martín et al. (2014) pay to habituation, a widespread behavioral phenomenon where continued exposure to a stimulus decreases behavioral response to that stimulus. Martín et al. (2014) argue that a virtue of their model is that it simulates habituation in a way that generally agrees with experimental work on habituation. Given the argumentative context in which I am using Martín et al. (2014) as a case study, it is interesting to note that habituation is one of the behavioral phenomena Lorenz argued his hydraulic model of instinctive behavior could account for (Burghardt 1973, p. 332–333).

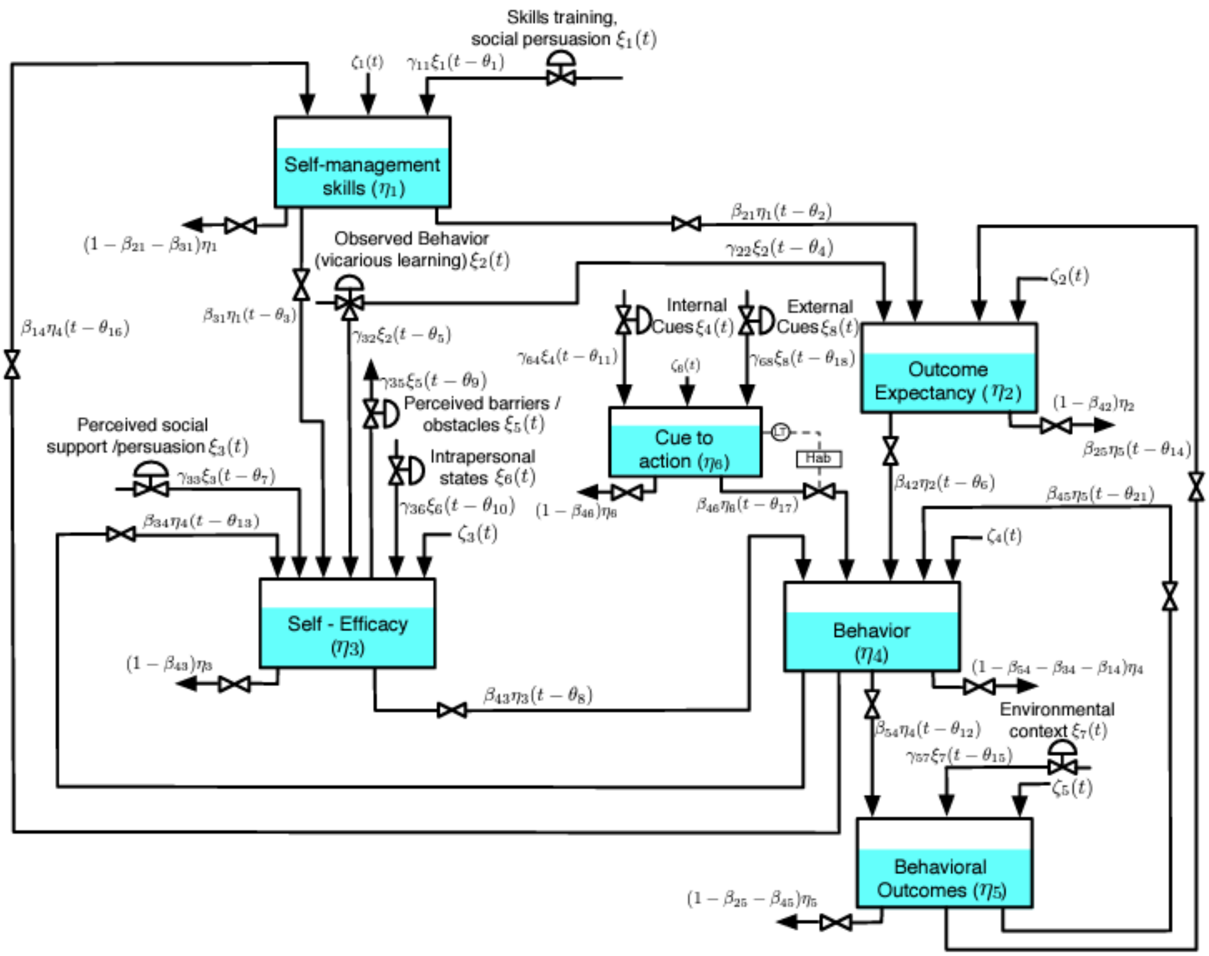

Figure 3: Martín et al.'s (2014) hydraulic dynamical systems model of social cognitive theory

However, Martín et al.'s (2014) use of motivational instinct is much broader than Wolf et al.'s (2018). In Lorenz's and Wolf et al.'s (2018) hydraulic models, inputs are sensory, outputs are behavioral, and everything in between is analogous to some endogenous physiological process. Martín et al.'s (2014) model, by contrast, takes more sophisticated cognitive inputs like "perceived social support and verbal persuasion" and "skills training" in addition to sensory inputs. Behavior is also just one category of outputs for Martín et al.'s (2014) model, the others can be beliefs about "self-efficacy" or "outcome expectancies". These output categories are represented by tanks full of liquid, and variations in output type are represented by variations in the amount of liquid in the tank (e.g. if the behavior tank is almost full, the agent is performing some behavior relatively frequently or for a relatively long duration.) Feedback loops form because output tanks are connected to other output tanks via pipes with well specified inflow and outflow resistances. As time progresses and the fluid flows, observations are made, beliefs are fixed, and behaviors are performed. The liquid in Martín et al.'s (2014) model is difficult to interpret because it represents relationships between psychological and behavioral events, whatever the nature of those relationships is.

The point of this section was to demonstrate that behavioral scientists use instinct concepts in different ways, that they have combined instinct and learning in diverse ways to create causal physiological accounts of behavior, and that they continue to do so.

4. Instinct Concepts Facilitate Abductive Inferences Between Causal Physiological Properties of Behavior

In this section, I suggest a positive epistemic role for instinct concepts. In section 2, I considered two critiques of instinct. The first only applied to investigations aimed at the developmental dimension of behavior, so my present suggestion about the value of instinct concepts for investigating the causal physiological dimension of behavior is outside the scope of that critique. The second critique concerned the sorts of inferences instinct concepts license. The idea was that when scientists deem a behavior instinctive because the behavior possesses some instinctive property, that judgement erroneously licenses further deductive inferences that the behavior also possesses other instinctive properties. I argue that instead of *deducing* that a behavior possesses one instinctive property based on the established presence of another instinctive property, scientists can *abduce* the same the conclusion.

In the context of scientific investigation, deduction and abduction lead to different outcomes. Deduction leads one to fix a belief as certain so long as prior premises are sound. Abduction leads one to form a hypothesis that explains some prior state of affairs (Peirce 1960; Niiniluoto 1999; Schurz 2008). Abductive inferences do not lead to certainty. Rather, they lead to further experiments so long as the abductive explanation has implications that are amenable to scientific inquiry. One of the ambiguities of abductive reasoning is how people determine what explanations are better than others in practice and how people ought to make such judgements in principle. I suggest instinct concepts help guide scientists through the uncertainty of using abductive reasoning to propose explanatory hypotheses. They do this by offering scientists a collection of causal physiological properties that, taken together, constitute a cloud of loosely allied commitments about behavior. Instinct concepts constrain the possibility space of abductive explanations by promoting explanations that realize instinctive properties and discounting explanations that conflict with instinctive properties.

This talk of instinct concepts containing multiple properties is reminiscent of Boyd's (1991, 1999) homeostatic property cluster theory of natural kinds. However, the positive role I have outlined for instinct concepts does not depend on instinct being a natural kind. Indeed, given the historical pragmatic approach I take in this paper, the ontological status of instinct concepts is not a pressing concern. It is more interesting to examine instinct concepts as tools for scientific inquiry that develop over time as scientists put them to new uses. The reason for treating different concepts and properties as somehow allied under the banner of "instinct" is historical and practice-based. These concepts and their attendant properties seem to come in packs that move together across research communities and time. I treat instincts as umbrella concepts, a term that I take to be agnostic about whether something is natural kind.

Many defenders of instinct have sought to define instinct in terms of more specific behavioral properties (Cowie 1999; Marler 2004; Pinker 2004; Samuels 2004, 2007; Collins 2005; Ariews 2007; Birch 2009; Perovic & Radenovic 2011; Margolis and Laurence 2013; Tabery 2014; Khalidi 2016; Cofnas 2017; Cain 2021; Ritchie 2021). The problem is that such attempts tend to retain properties characterizing the developmental dimension of behavior (e.g. developmental-fixedness). In other words, the problem is that instinct defenders keep trying to rescue a developmental notion of instinct.

So, abandoning developmental notions of instinct, what sorts of properties characterize causal physiological instinct concepts? I started section 3 by highlighting two different causal physiological notions of instinct: a motivational or drive-based conception of instinct and a reflex-based conception of instinct. To give a preliminary outline of the properties characterizing motivational and reflex-based instinct concepts, I draw on section 3's case studies and the properties that appear to characterize instinct concepts in those studies. But first, I highlight some causal physiological properties that defenders of instinct have already used to characterize instinct.

For example, Margolis and Laurence (2013, p. 697, emphasis in original) argue that "To understand what nativism consists in, you have to see how nativism contrasts with empiricism. You have to focus on what empiricists and nativists actually disagree about. This is primarily *the character of the psychological systems that underlie the acquisition of psychological traits*." More specifically, Margolis and Laurence (2013, p. 696) argue that

> "Nativists explain the richness and variety of psychological outcomes in large part by reference to the diversity of the psychological acquisition systems […] Moreover, nativists suppose that many of these systems are domain-specific systems that are not

acquired on the basis of more fundamental domain-general psychological systems. So for nativists, the acquisition of all psychological traits ultimately depends not only on domain general psychological systems, but also on a large number of domain-specific psychological systems whose acquisition cannot in turn be explained in psychological terms."

The above quote focuses on a developmental property, the property of not being "acquired on the basis of more fundamental domain-general psychological systems." But it also references a causal physiological property, the property of being "domain-specific". As a first definitional pass, behavioral systems have more domain-specificity when they are responsible for achieving fewer functions, and vice versa. Domain specificity also seems to involve the idea that the functions are somehow of a piece—that they are not a collection of haphazard and incongruous functions. For example, the functions of a highly domain-specific mechanism may be jointly responsible for realizing some higher-level capacity, in which case this capacity is the "domain" that the mechanism's functions are "specific" to. The number of functions a mechanism is supposed to perform can be a causal physiological property in the sense that it provides normative valence to the mechanism's causal physiological dynamics, and this valence has implications for investigating the causal physiological dimensions of that mechanism. For example, if scientists believe a mechanism is domain-general, but they have only elucidated the causal pathways for a few specific functions, then it makes sense to design exploratory experiments aimed at uncovering more causal pathways for more functions.[8]

[8] Those who think all biological function concepts are historical may take issue with my definition of domain-specificity on the grounds that by appealing to functions, I am covertly appealing to history and development. I do not have space to defend my appeal to biological function here, but as shown by Dhein (2020), the way experimental behavior scientists ascribe

Cowie (1998) articulates a property similar to domain-specificity when she argues that one of the subterranean debates motivating the nativism versus empiricism debate in behavior is the debate over whether animals possess special purpose learning faculties or domain general learning faculties. Cowie argues that nativists promote special purpose learning faculties and empiricist promote domain general learning faculties. The relationship between special purpose learning faculties and domain specificity is clear: special purpose learning faculties are more domain-specific than general learning faculties.

Finally, and in a similar vein, Khalidi (2016) argues that the cluster of properties surrounding innate cognitive capacities are causally linked in a way that suggests they constitute a natural kind. Some of the properties Khalidi (2016, p. 321) includes in his cluster have obvious causal, non-developmental interpretations:

- "**Triggering** (or more properly, triggerability). Can be acquired in conditions of relative informational impoverishment.
- **Informational encapsulation.** Insulated from other cognitive content, functions independently of other cognitive systems.
- **Cognitive impenetrability.** Resists modification by other cognitive capacities."

The last two properties, "informational encapsulation" and "cognitive impenetrability", have straightforward causal physiological interpretations that seem allied with Margolis and Laurence's (2013) domain-specificity and Cowie's (1998) special purpose faculties. The first property, however, speaks to how cognitive capacities are "acquired". To make triggerability more causal, one must modify it so that it does not refer to the acquisition of traits over the time

biological function to physiological mechanisms requires minimal assumptions about the history of those mechanisms. See García-Valdecasas and Deacon (2024) for a similar distinction.

frame of an animal's lifecycle but to the immediate preconditions for the performance of a behavior. A causal notion of triggerability belongs to reflex-based notions of instinct exemplified by Craig's (1918) consummatory reaction. The triggered behavior is automatic in the counterfactual sense that once triggered, the animal cannot decide to do otherwise. It is also automatic in the sense that the animal need not be aware of the adaptive purpose fulfilled by the triggered behavior. Finally, instinctive behaviors may also be triggered in the sense that they do not require further inputs to properly unfold; the initial trigger is sufficient.

Scientists in the Wolf et al. (2018) case study from section 3 appear to make abductive inferences between some of the causal physiological properties articulated above. After performing navigation experiments on foraging ants, Wolf et al. (2018) built a Lorenzian model of memory integration that encapsulates their new hypotheses (Figure 2). They explicitly state five hypotheses embedded in the model:

I) Separate odometer memories charge independently and in parallel.

II) The two odometers interact with different relative weights.

III) Odometer memories maintain their charged value if the sensory mechanisms associated with those memories are not active.

IV) Different odometer memories discharge separately.

V) When an ant depletes one of her odometer memories, it triggers nest search behavior.

Hypotheses I, III, and IV are applied instances of the causally-insulated and functionally-specific properties. That is to say, the properties of causally-insulated and functionally-specific have guided Wolf et al. (2018) through the possibility space of hypotheses they could have abducted to explain their experimental results. Hypothesis V shows the mark of the "triggerability" causal property just discussed. Only hypothesis II falls outside the property list. It posits a causal interaction between the two modules.

Returning to Brigandt's (2010) three part analysis of scientific concepts, Brigandt characterizes scientific concepts in terms of (1) reference, (2) inferential role, and (3) the epistemic goal pursued with the concept's use. In Wolf et al. (2018), the referents of the instinct properties are unknown neurophysiological mechanisms controlling the way two measures of distance traveled interact to guide navigation. The abductive role I have outlined for instinct concepts means that the referent of such concepts may often be largely unknown. This is in line with Feest's (2017) argument that experimental work in the cognitive sciences is often directed at describing and exploring an object of research. Similarly, in defending a causal notion of instinct, Burghardt and Bowers (2017, p. 338) write that "Lorenz recognized that instinct set the stage for analytic study, rather than constituting an explanation." Wolf et al.'s (2018) instinct-influenced hydraulic model of memory suggests new experiments to test new hypotheses.

Clarifying the connection between a motivational conception of instinct, Lorenz's use of a fluid analogy to model that concept, and Wolf et al.'s (2018) and Martín et al.'s (2014) extension of that fluid analogy requires further work. What, for instance, are the practical consequences of thinking about behavior in terms of energy or liquid flowing through channels? One possible answer is that this concept of instinct biases scientists away from hypotheses that rely on cognitive psychological concepts like "belief" and toward hypotheses that rely on physical concepts like "force". Wolf et al. (2018) take this physical, non-psychological framing deep into the learning process so that it mediates the use of memories in ants. Tolman, by contrast, pushed the motivational notion of instinct far into the background. Tolman's explanations of behavior mainly invoke cognitive psychological concepts, but those concepts cannot apply to all the constituent parts of a behavioral system (e.g. molecules do not form beliefs). So, he uses motivational instinct to provide a vague, non-psychological, goal-directed foundation for psychological concepts. Martín et al.'s (2014) hydraulic model of social cognitive theory is the product of control systems engineering and classic learning psychology (Hekler et al. 2018). It is easy to imagine the hydraulic aspect of the model coming from control systems engineering. That field has developed powerful concepts and methods for predicting and optimizing the dynamics of physical systems. By modeling psychological phenomena as physical, scientists hope to expand the power of physical concepts into the domain of the psychological.

Cognitive science is a big tent that includes conflicting research programs and theoretical commitments (Allen 2017). Instinct properties are one way of dealing with the challenge of formulating hypotheses in such an unconstrained space. And despite the superficial a priori triviality of the instinct versus learning dichotomy, scientists have formulated significantly different notions of instinct and learning over time. They have also combined those different conceptions in creative ways. Given this paper's use of case studies to suggest that instinct concepts have epistemic value for causal physiological studies, it's premature to abandon instinct concepts on the grounds of 20th century arguments against developmental notions of instinct.

5. Conclusion

The dichotomies of "nature versus nurture", "nativism versus empiricism", "inheritance versus environment", and "innate versus acquired" have been a common theme throughout the history of the behavioral sciences. I have argued that 20th century interactionist arguments against instinct only work against developmental notions of instinct. I have also argued that the inferential argument against instinct is restricted to a particular pattern of scientific reasoning. To demonstrate the combinatorial flexibility of instinct and learning concepts, I examined multiple case studies from the 20th and 21st centuries. To sketch a positive epistemic role for instinct concepts in the behavioral sciences, I argued that instinct concepts can be defined in terms of causal physiological properties, not just developmental properties. I then suggested that instinct concepts facilitate abductive inferences between causal physiological properties to guide hypothesis formation.

References

(Eds.) Aronson, Lester, Ethel Tobach, Daniel Lehrman, and Jay Rosenblatt. Development and evolution of behavior: Essays in memory of T. C. Schneirla. Freeman 1970

Allen C (2017) On (not) Defining Cognition. *Synthese* 194(11): 4233–4249

Arabatzis T (2012) Experimentation and the meaning of scientific concepts. Scientific concepts and investigative practice, 149-166

Ariew A (1986) Innateness. In Philosophy of biology, pp. 567-584. North-Holland, 2007

Bandura A (1986) Social foundations of thought and action. Englewood Cliffs, NJ: Prentice Hall

Barlow GW (1991) Nature-nurture and the debates surrounding ethology and sociobiology. American Zoologist, 31(2), 286-296

Beach FA (1955) The descent of instinct. Psychological Review, 62(6), 401

Beer CG (1975) Was Professor Lehrman an ethologist? Animal Behaviour 23 957–64

Beer CG (2020) Niko Tinbergen and questions of instinct. Animal Behaviour, 164, 261-265

Bernard LL (1921) The Misuse of Instinct in the Social Sciences. Psychological Review 28(2):96-119

Birch J (2009) Irretrievably confused? Innateness in explanatory context. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 40(4):296-301

Bitterman, M. E., Menzel, R., Fietz, A., & Schäfer, S. (1983) Classical conditioning of proboscis extension in honeybees (Apis mellifera). Journal of Comparative Psychology, 97(2):107

Boakes R (1984) From Darwin to behaviourism: Psychology and the minds of animals. CUP Archive.

Boon M (2012) Scientific Concepts in the Engineering Sciences. Scientific concepts and investigative practice, 219-243

Boyd R (1991) Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. Philosophical Studies 61:127–148

––– (1999) Homeostasis, Species, and Higher Taxa. in R. Wilson (ed.), Species: New Interdisciplinary Essays, Cambridge, MA: MIT Press: 141–186.

Brandom R (1994) Making it explicit: Reasoning, representing, and discursive commitment. Harvard university press.

Bregy, P., Sommer, S., & Wehner, R (2008) Nest-mark orientation versus vector navigation in desert ants. Journal of Experimental Biology 211(12): 1868–1873

Brigandt I (2005) The instinct concept of the early Konrad Lorenz. Journal of the History of Biology 38:571-608

Brigandt I (2010) The epistemic goal of a concept: accounting for the rationality of semantic change and variation. Synthese, 177(1):19-40

Brigandt I (2012) The dynamics of scientific concepts: The relevance of epistemic aims and values. Scientific concepts and investigative practice, 3, 75-103

Burgess N (2014) The 2014 Nobel Prize in Physiology or Medicine: a spatial model for cognitive neuroscience. Neuron, 84(6):1120-1125

Burghardt GM (1973) Instinct and innate behavior: Toward an ethological psychology. In J. A. Nevin & G. S. Reynolds, The Study of Behavior: Learning, Motivation, Emotion, and Instinct. Scott, Foresman.

Burghardt GM (2009) Darwin's legacy to comparative psychology and ethology. American Psychologist, 64(2):102

Burghardt GM (2020) Insights found in century-old writings on animal behaviour and some cautions for today. Animal Behaviour 164:241-249

Burghardt GM, Bowers RI (2017) From instinct to behavior systems: An integrated approach to ethological psychology. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), APA handbook of comparative psychology: Basic concepts, methods, neural substrate, and behavior (pp. 333–364). American Psychological Association

Burghardt GM, Burkhardt RW (2018) Wallace Craig's Appetites and aversions as constituents of instincts: A centennial appreciation. Journal of Comparative Psychology 132(4):361

Burkhardt RW (2005). *Patterns of behavior: Konrad Lorenz, Niko Tinbergen, and the founding of ethology*. University of Chicago Press

Cain MJ (2021) Innateness and Cognition. Routledge

Cheng K, & Freas, CA (2015) Path integration, views, search, and matched filters: the contributions of Rüdiger Wehner to the study of orientation and navigation. Journal of Comparative Physiology A201:517-532

Cofnas N (2017) Innateness as genetic adaptation: Lorenz redivivus (and revised). Biology & Philosophy 32(4):559-580

Collins J (2005) Nativism: In defense of a biological understanding. Philosophical Psychology 18(2):157-177.

Cowie F (1999) What's within?: Nativism reconsidered. Oxford University Press

Craig W (1918) Appetites and aversions as constituents of instincts. The Biological Bulletin 34(2):91-107

Dewsbury D (1992) Comparative psychology and ethology: A reassessment. Amer. Psych. 47(2):208–15

Dewsbury D (2002) The Chicago Five: A family group of integrative psychobiologists. History of psychology 5(1):16.

Dhein K. "What makes neurophysiology meaningful? Semantic content ascriptions in insect navigation research." Biology & Philosophy 35, no. 5 (2020): 52.

Dhein K. "Karl von Frisch and the Discipline of Ethology." Journal of the History of Biology 54, no. 4 (2021): 739-767.

Dhein K. "The cognitive map debate in insects: A historical perspective on what is at stake." Studies in History and Philosophy of Science 98 (2023): 62-79.

Dunlap K. (1919) Are there any instincts?. The Journal of Abnormal Psychology 14(5): 307.

Eibl-Eibesfeldt I (1961) The interaction of unlearned behaviour patterns and learning in mammals. In A. Fessard, R. W. Gerard, & J. Konorski (Eds.), Brain mechanisms and learning: A symposium (pp. 53-73). Oxford: Blackwell.

Eibl-Eibesfeldt I (1997) Human Ethology: Origins and Prospects of a New Discipline." In *New Aspects of Human* Ethology. Eds. Schmitt, Alain, Klaus Atzwanger, Karl Grammer, Katrin Schäfer. p. 1–24. Plenum Press, New York.

Ewert, JP (1970) Neural mechanisms of prey-catching and avoidance behavior in the toad (Bufo bufo L.). Brain Behavior and Evolution 3(1-4): 36-56.

Ewert, Jörg-Peter, Harald Burghagen, and Evelyn Schürg-Pfeiffer (1983) Neuroethological analysis of the innate releasing mechanism for prey-catching behavior in toads. Advances in vertebrate neuroethology: 413-475.

Ewert, JP (1987) "Neuroethology of releasing mechanisms: prey-catching in toads." Behavioral and Brain Sciences 10, no. 3 (1987): 337-368.

Feest U (2005) Giving up instincts in psychology--or not? Passauer Schriften zur Psychologiegeschichte 13: 242-259.

Feest U (2010) Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology. Spontaneous Generations: A Journal for the History and Philosophy of Science 4(1): 173-190.

Feest U (2012) Exploratory experiments, concept formation, and theory construction in psychology. Scientific concepts and investigative practice 3:167-189.

Feest U (2017) Phenomena and objects of research in the cognitive and behavioral sciences. *Philosophy of Science* 84(5): 1165-1176.

García-Valdecasas, M., & Deacon, T. W. (2024) Biological functions are causes, not effects: A critique of selected effects theories. Studies in History and Philosophy of Science, 103, 20-28.

Griffiths, PE, and RD Gray (1994) "Developmental systems and evolutionary explanation." The Journal of philosophy 91(6): 277-304.

Griffiths, PE (2002) What is innateness? *The Monist* 85(1): 70-85.

Griffiths, PE (2004) Instinct in the '50s: The British Reception of Konrad Lorenz's Theory of Instinctive Behavior Biology and Philosophy 19(4): 609–631.

Griffiths, PE (2008) Ethology, Sociobiology, and Evolutionary Psychology in A Companion to the Philosophy of Biology, ed. Sahotra Sarkar and Anya Plutynski (New Jersey: John Wiley & Sons, 2008), 393–414.

Griffiths, P, E Machery, and S Linquist (2009) "The vernacular concept of innateness." Mind & Language 24(5): 605-630.

Griffiths, P, and S Linquist (2022) The distinction between innate and acquired characteristics. The Stanford Encyclopedia of Philosophy (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.)

Hebb, DO (1953) Heredity and environment in mammalian behaviour. British Journal of animal behaviour

Hinde, RA (1966) *Animal Behaviour: A Synthesis of Ethology and Comparative Psychology*. New York: McGraw-Hill.

Hirsch, J., Lindley, R. H., & Tolman, E. C. (1955). “An experimental test of an alleged innate sign stimulus.” Journal of Comparative and Physiological Psychology, 48(4), 278–280.

Hoinville, T., & Wehner, R. (2018). Optimal multiguidance integration in insect navigation. *Proceedings of the National Academy of Sciences*, *115*(11), 2824-2829.

Jamieson, A., & Radick, G. (2013). Putting Mendel in his place: How curriculum reform in genetics and counterfactual history of science can work together. In The philosophy of biology: A companion for educators (pp. 577-595). Dordrecht: Springer Netherlands.

Khalidi, MA. Innateness as a natural cognitive kind. Philosophical Psychology 29(3): 319-333.

Kindi, V (2012) "Concept as vessel and concept as use." Scientific concepts and investigative practice: 23-46.

Knobe, J, and Richard S (2013) "Thinking like a scientist: Innateness as a case study." Cognition 126(1): 72-86.

Kuo, ZY (1921) Giving up instincts in psychology. Journal of Philosophy 18:645-664.

Radick, G (2005) "Other histories, other biologies." Royal Institute of Philosophy Supplements 56: 21-47.

Radick, G (2016) "Presidential address: Experimenting with the scientific past." The British Journal for the History of Science 49(2): 153-172.

Robinson, GE (2004) "Beyond nature and nurture." Science 304(5669): 397-399.

Rosenblatt, JS (1995) "Daniel S. Lehrman 1919–1972." In: National Academy of Sciences (ed.): Biographical Memoirs, p. 225–245

Lehrman, DS (1953) A critique of Konrad Lorenz's theory of instinctive behavior. The Quarterly review of biology 28(4): 337-363.

Lehrman, DS (1970) "Semantic and conceptual issues in the nature-nurture problem." LR Aronson et al.(Eds.), Development and Evolution of Behavior. Essays in memory of TC Schneirla, San Francisco (WH Freeman and Company), pp. 17-52.

Lewontin, RC (1979) "Sociobiology as an adaptationist program." Behavioral science 24(1): 5-14.

Linquist, S (2018) "The conceptual critique of innateness." *Philosophy Compass* 13(5): e12492.

Lorenz, K. (1937). Über die Bildung des Instinktbegriffes. *Die Naturwissenschaften* 25, 289–300, 307–318, 324–331.

Lorenz, K. (1950). The comparative method in studying innate behaviour patterns. Symposia of the Society for Experimental Biology 4, 221–268.

Mameli, M (2008) "On innateness: The clutter hypothesis and the cluster hypothesis." The Journal of philosophy 105(12): 719-736.

Mameli, M. and Bateson, P (2006) Innateness and the sciences. Biology and Philosophy, 21, pp.155-188.

Mameli, M and P Bateson (2011) "An evaluation of the concept of innateness." Philosophical Transactions of the Royal Society B: Biological Sciences 366(1563): 436-443.

Margolis, E and S Laurence (2013) "In defense of nativism." Philosophical Studies 165 (2013): 693-718.

Marler, P (2004) "Innateness and the instinct to learn." Anais da Academia Brasileira de Ciências 76: 189-200.

Martín, Cesar A., Daniel E. Rivera, William T. Riley, Eric B. Hekler, Matthew P. Buman, Marc A. Adams, and Abby C. King. "A dynamical systems model of social cognitive theory." In *2014 American Control Conference*, pp. 2407-2412. IEEE, 2014.

Menzel, R (2019a) "The waggle dance as an intended flight: a cognitive perspective." Insects 10, (12): 424.

Menzel, R (2019b) “Search Strategies of Intentionality in the Honeybee Brain.” In *The Oxford Handbook of Invertebrate Neurobiology*. p. 663–83.

Menzel, R., Brandt, R., Gumbert, A., Komischke, B., & Kunze, J. (2000). Two spatial memories for honeybee navigation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1447), 961-968.

Menzel, R. (2019). The Waggle Dance as an Intended Flight: A Cognitive Perspective. *Insects*, 10(12), 424.

Niiniluoto, I (1999) Defending abduction. Philosophy of science, 66(S3), S436-S451.

Oyama, S (2000) The ontogeny of information: Developmental systems and evolution. Duke university press

Patton, L 2022 Organic Memory and the Perils of Perigenesis: The Helmholtz-Hering Debate. In *Mechanism, Life and Mind in Modern Natural Philosophy* 345–362. Cham: Springer International Publishing.

Peirce, CS (1878) “How to Make Our Ideas Clear.” Popular Science Monthly, 12, 286–302.

Peirce, CS (1960) Collected Papers of Charles Sanders Peirce: Pragmatism and pragmaticism and Scientific metaphysics. Cambridge: Belknap Press.

Perovic, S and L Radenovic (2011) "Fine-tuning nativism: the ‘nurtured nature’and innate cognitive structures." Phenomenology and the Cognitive Sciences 10: 399-417.

Pinker, S (2004) "Why nature & nurture won't go away." Daedalus 133(4): 5-17.

Ritchie, JB (2021) What’s wrong with the minimal conception of innateness in cognitive science?. Synthese, 199(Suppl 1), 159-176.

Ronacher B (2019) Innate releasing mechanisms and fixed action patterns: basic ethological concepts as drivers for neuroethological studies on acoustic communication in Orthoptera. Journal of Comparative Physiology A, 205(1), 33-50.

Samuels R (2004) Innateness in cognitive science. Trends in cognitive sciences, 8(3), 136-141.

Samuels, R (2007) Is innateness a confused concept. The innate mind, 3, 17-36.

Schurz, G (2008) Patterns of abduction. Synthese, 164, 201-234.

Sellars, W (1968). Science and metaphysics: variations on Kantian themes. New York,: Humanities P..

Senchuk, DM (1991) "Against instinct: From biology to philosophical psychology."

Shea, N (2012) Genetic representation explains the cluster of innateness-related properties. Mind & Language, 27(4), 466-493.

Skinner, BF (1965) Science and human behavior (No. 92904). Simon and Schuster.

Smith, ET (2020) Examining tensions in the past and present uses of concepts. *Studies in History and Philosophy of Science Part A*, 84, 84-94.

Steinle, F (2012) Goals and fates of concepts: The case of magnetic poles. Scientific concepts and investigative practice, 105-126.

Stich, S (1975) The idea of innateness. In S. Stich (Ed.), Innate ideas. Los Angeles: University of California Press.

Tabery, James. Beyond versus: The struggle to understand the interaction of nature and nurture. mit Press, 2014.

Tinbergen, N (1963) On aims and methods of ethology. *Zeitschrift für tierpsychologie*, 20(4), 410-433.

Tolman, EC (1932) *Purposive Behavior in Animals and Men*. University of California Press.

Tolman, EC (1948). "Cognitive maps in rats and men." Psychological review 55(4): 189.

Vicedo, M (2023) Beyond the instinct debate: Daniel Lehrman’s contributions to animal behavior studies. Journal of the History of Biology, 1-34.

Wehner, R., & Wehner, S. (1990). Insect navigation: use of maps or Ariadne's thread? *Ethology Ecology & Evolution*, *2*(1), 27-48.

Wehner, R., Hoinville, T., Cruse, H., & Cheng, K. (2016). Steering intermediate courses: Desert ants combine information from various navigational routines. Journal of Comparative Physiology, 202(7), 459–472.

Wittgenstein, L (1953) Philosophical Investigations, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.

Wolf, H., Wittlinger, M., & Pfeffer, S. E. (2018). Two distance memories in desert ants—Modes of interaction. PLoS One, 13(10), e0204664.