



---

## Accepting the Povinelli-Henley Challenge

Nancy A. Salay

Department of Philosophy and School of Computing, Queen's University

Email: [salay@queensu.ca](mailto:salay@queensu.ca)

**Citation** – Salay, N. A. (2022). Accepting the Povinelli-Henley challenge. *Animal Behavior and Cognition*, 9(2), 239-256. <https://doi.org/10.26451/abc.09.02.08.2022>

**Abstract** – In the recent twenty-year retrospective issue of *Animal Behavior and Cognition*, Povinelli and Henley (2020) argue that a host of comparative studies into “complex cognition” suffer, fatally, from a theoretical confusion. To rectify the problem, they issue the following challenge: alongside specifications of the higher-order capacity to be tested, provide hypotheses of the mechanism(s) *necessary* to implement it. They spearhead this effort with a discussion of how the Relational Reinterpretation Hypothesis (RRH) provides just such an account. In the first part of the paper, I argue that RRH is neither necessary nor sufficient to explain the second-order behavior in question. In part two, I describe an alternative hypothesis, externalism, that does sufficiently account for it. Further, it opens new avenues of comparative research.

**Keywords** – Representation, Externalism, Embodied, Causal, Cognition

---

In the recent twenty-year retrospective issue of *Animal Behavior and Cognition*, Povinelli and Henley (2020) argue, convincingly, that cleverly crafted tool-use problems are demonstrably incapable of providing evidence of higher-order capacities: “... no experimental approach that presents physical (or virtual) objects to animals (as tools or otherwise), and then measures whether or how quickly they demonstrate mastery of objective causal facts about them, will ever allow for strong inferences about the presence of higher-order relational reasoning” (Povinelli & Henley, 2020, p. 408). Recognising this problem is one step closer to solving it, they maintain, because it forces cognitively-minded researchers to clarify their theoretical commitments. To this end, they issue the following challenge to the community: before developing empirical tests, provide a functional specification of the higher-order capacity to be tested and an algorithmic specification of the mechanisms *necessary* to implement it. Following that, identify exemplifying human behavior that can serve as a benchmark for this specified comparative work.

This call for theoretical grounding is sound counsel; unfortunately, they do not follow their own advice. Povinelli and Henley offer the Relational Reinterpretation Hypothesis (RRH) as a partial algorithmic specification of certain higher-order capacities: “... only human animals possess the representational processes necessary for systematically reinterpreting first-order perceptual relations in terms of higher-order, role-governed relational structures akin to those found in a physical symbol system” (Penn et al., 2008a, p. 111). But, as I shall argue here, they fail to show that RRH is necessary to explain those capacities. Given this, the directive that (in phase three) experiments must focus on “... understanding how higher-order reasoning modulates adult *human* behavior in ecologically relevant tasks” (Povinelli & Henley, 2020, p. 418, emphasis added) belies a double-standard. In the absence of an argument that biologically fundamental representational mechanisms are necessary for “higher-order reasoning,” the exhortation to use human behavior as an experimental benchmark is as unjustified as the second-order

attributions are in the comparative cases: humans “could have learned the relations in question (how to use the tools) without the mediating influence of ... [internal] higher-order representations ...” (ibid., p. 404).

For RRH to be a viable hypothesis, an algorithmic account of it will need to address a range of problems that plague all reductive representationalist accounts of cognition: In what sense do lower-level mechanisms *function* as representations; How are lower-level representational mechanisms related to higher level mechanisms that also function as representations; What determines the content of these representation-bearing processes? There are good reasons for wondering whether solutions to these are ever forthcoming (Chemero, 2009; Clark & Toribio, 1994; Dreyfus, 2007; Freeman, 2000; Keijzer, 1998; Van Gelder, 1995) and so meeting the phase one challenge, at least for RRH, is a tall order.

Part of the problem is that Povinelli and Henley incorrectly see the field of pre-theoretical commitments as dividing between those working within the representational paradigm and those working outside of it. Thus, they dismiss considerations raised by Fragaszy and Managalam (2020)<sup>1</sup> on the grounds that “... these alternatives are even less friendly toward the current *experimental project* of trying to demonstrate higher-order thinking in animals than ours” (Povinelli & Henley, 2020, p. 417). But this oversimplifies the current theoretical landscape. Some of the most interesting work in contemporary cognitive science is emerging at the intersection of cognitivist and embodied/enactive insights about cognition in which representation, embodiment, and environment all play a pivotal role in cognition. From this vantage point, the problem with RRH lies not with its functional level representational commitments, but with the internalist assumption that an algorithmic specification of them necessarily involves a reduction to biologically fundamental capacities of individual humans. An alternative, what I will be calling “externalism,” to highlight that it is RRH’s commitment to reductionism rather than to representationalism that is the point of contention, holds that second-order cognitive behavior does not necessarily entail a second-order *capacity* at all: sophisticated cognitive behavior could be the result of simple, first-order cognitive processes buttressed by cognitively expanding social practices, e.g., language. If externalism is right, the benchmark of human cognitive capacity that will serve as the ground of comparison for future comparative research will look very different to that imagined by Povinelli and Henley from their RRH vantage point. Furthermore, the cognitive gap between humans and non-human animals will narrow, thus absolving us of the need for questionable evolutionary speculation.<sup>2</sup>

After taking a closer look at RRH and examining the (Penn et al., 2008a) response to a version of externalism, I elaborate on how, properly understood, an externalist account of second-order, relational judgements side-steps the need to appeal to higher-order *capacities*. If externalism is right, or even a viable theoretical alternative, Povinelli and Henley’s diagnosis of what ails cognitivist comparative psychology is not nearly as dire as it first appears.

### Terminological Clarifications

To begin, some terminological clarifications are in order, since key terms such as “representation” and “function” are used in a confusingly wide variety of ways in the cognitive sciences. As this is an examination of RRH, I will follow Penn et al. in usage except where an alternative is warranted in the interest of clarity.

In Penn et al. (2008a), RRH is presented in terms of David Marr’s (1982) levels of analysis for information systems, though they make use of only two of these levels, what they call the functional and representational levels: “The RRH provides a representational-level account of the functional differences between human and animal cognition by distinguishing between the cognitive requirements for first-order, perceptually-based relational reasoning versus higher-order, structural, role governed relational reasoning” (p. 110). That they are appealing to Marr is clear from a comment to their commentators: “The problem here is not an “anomaly” in our logic—the problem is that Wasserman does not acknowledge the difference between a functional- and a representational-level of analysis (Marr 1982)” (Penn et al., 2008b, p. 158).

<sup>1</sup> As well as anyone taking an anti-representational stance. See Povinelli and Henley (2020, footnote 12).

<sup>2</sup> As in, for example, Berwick and Chomsky (2015).

Wasserman<sup>3</sup> can be excused for not realizing this, however, since the labels they use for the various levels of analysis do not correspond to Marr's own usage. For him the highest level of description, the behavior to be explained, is "computational" — what Penn et al. call the "functional" level of analysis — and its formal specification is "algorithmic/representational" — what Penn et al. call the "representational" level of analysis. Though I will follow them in their use of the "functional" label, it seems prudent to adopt Marr's preferred term, "algorithmic," to refer to the formal specification of the functional description since the term "representational" is frequently used in the cognitive context in many more ways than to refer to these levels of analysis.

Having established the stipulative usage of these key terms, the central critical observation of Povinelli and Henley (2020) can be stated in terms of them: in comparative research functional level descriptions of behavior are confused with algorithmic level speculations, for example, about "the kinds of causal information animals rely upon when using tools" (p. 393). But these are not warranted since, as they rightly point out, necessity relations between levels of explanation extend from the top-down, not the bottom-up: "... strong evidence for higher-order representations in a cognitive system is strong evidence for first-order, perceptual-based representations, but not vice versa" (ibid., p. 394).<sup>4</sup> In other words, so long as we can account for functional level descriptions of behavior — complex tool use — by first-order relations, there is no warrant for positing higher-order representations such as <gravity>. For Povinelli and Henley (2020), a "first-order, perceptual-based representation" (p. 394) is a simple mental construct of an object currently being perceived: the cat, eyes trained on the mouse, perceives it. This is a first-order relation because the relevant relatum is the object of perception itself, whether we analyse that to be an internal representation or an external object. Humans, however, are also capable of recognizing relations that hold *between* objects. While crossing the road, I notice an approaching car. I judge that the speed and distance of the car relative to myself and my current trajectory raises no cause for concern and I continue crossing at my normal pace. In making this judgement, I have compared two relations, the relation between me and the car and the relation between me and the other side of the road. The distances I am comparing are relations *between* the physical objects in my environment, that is, second-order relations. The received view is that the ability to make inferences using these higher-order relations sets human cognitive capacity apart from that of non-human animals. How to test for whether non-human animals make inferences that depend on these second-order relations is the question under discussion. With these clarifications in hand, I turn now to the details of RRH.

### Unpacking the Relational Reinterpretation Hypothesis

The Relational Reinterpretation Hypothesis (RRH) is the claim that the computational representational theory of mind is literally true, that the human cognitive system is a physical symbol system. If this were all there were to RRH, that is, if it were *only* a functional level account, there would be nothing to quibble about: humans who have developed in our modern cognitive niche, who have an array of cognitive tools at their disposal, really do behave, at least some of the time, in ways that implement symbol systems. Your ability to read and understand this paper is a case in point.

RRH is not only a functional level hypothesis, however; the implicit, and frequently explicit, assumption underwriting it is that our higher-order cognitive behavior is explained by representational, algorithmic level mechanisms that are implemented by our nervous systems. In other words, the capacity to make second-order judgements is biologically fundamental for humans: we "manifestly wield ... higher-order, structural, role-based, aspects of cognition that ... allow humans to redescribe perceptually-based categories and relations such that they can serve as the fodder for non-perceptually-based categories and relations involving constructs such as <weight>, <mass>, <shape>, <color>, <time>, <force>, <gravity>, <mental states> ..." (Povinelli & Henley, 2020, p. 393). The implicit assumption here is that

<sup>3</sup> In the style of *Behavioral and Brain Sciences* articles, the target article (Penn et al., 2008a), is followed by a discussion — commentaries from experts in the field and a rebuttal (Penn et al., 2008b). Here they are responding to Edward Wasserman's commentary (Wasserman, 2008).

<sup>4</sup> As I shall argue, *strong evidence for* does not imply *necessarily entails*, which is the conclusion they draw.

the theoretical posit of internal, second-order representational resources is necessary to explain the phenomenon: we literally *use* these internal constructs when reasoning about our world. But if, as I shall argue, alternative explanations explain the capacity without invoking these constructs, this critical assumption is undermined. Worse, as I shall argue, not only are internal representational constructs not necessary to account for higher-order cognitive behavior, they do not provide a sufficient account of it either.

A word needs to be said about the concept of representation<sup>5</sup> since it is peculiar and, as we have just seen, it plays such a central role in RRH. For something, internal or external, to function as a representation, it must serve, *in some sense*, as a stand-in for what it represents; otherwise, there is no reason to describe it as a representation at all. In other words, we describe something as a representation when it causes (at least some of) the effects its object would cause were it present.

Sentences are clear examples of representations.<sup>6</sup> Fred yells “Fire!” and Sally quickly exits the building. Here Sally responds to what Fred’s utterance represents, namely the presence of fire, in addition to its tone, pitch and other physical features. Semantically different exclamations will elicit different responses. Fred exclaims “Wow!” and Sally smiles at Fred expectantly. Naturally occurring sounds, in contrast, invoke responses to their physical properties only.<sup>7</sup> The door slams shut in the wind and, reflexively, Sally jumps.

The functional level phenomenon to be explained here is how Fred and Sally, and by extension, all symbol users, become responsive to representational properties at all. The RRH answer, as with internalist views generally, is to posit lower level relations that mirror the personal level ones. Sally responds to the meaning of Fred’s utterance by way of her internal representation of *<fire>*, which includes phenomenal as well as subpersonal elements, and it is this internal response — a cascade of neural, muscular, respiratory activity — that ultimately gets Sally moving out of the building. To see why such reductions fail to explain the phenomenon in question, we need to look more closely at how they work.

When we look at the way an animal’s sensory system develops over its experiences, we often notice correlations between patterns of macro-level activity and patterns of low-level activity; e.g., the presence of an object *kind* in an animal’s sensory field with an identifiable pattern of neural activity. As top-down observers of these correlations, from our spatio-temporally extended vantage point, that low-level activity can be described in representational terms.<sup>8</sup> Sometimes these constructs help us explain and predict the animal’s behavior — the frog flicked its tongue because it saw a fly, that is, its neural fly-tracker module was active. But as Povinelli and Henley (2020) rightly point out, “the fact that animals keep track of (and act upon) the spatial trajectories of unsupported objects, in no way implies that animals also represent *<gravity>*” (p. 394). Likewise, the fact that animals develop responses to object kinds, by way of the statistical regularities of their low-level features, in no way implies that animals also represent them. Simple machine learning classifiers are successful when their outputs align with the feature bundles they are designed to track. We might describe the weightings between the nodes of such systems as implicit representations of the natural kinds such systems identify, but there is no system-level *use* of these representations of the sort that marks the high-level cognitive activity we are seeking to understand, namely, planning, problem solving, and second-order reasoning. It is certainly tempting to view the network activity that cascades along these weighted associations as constituting that use: when one subpattern of activity triggers all associated subpatterns, it looks like it is functioning as an indicator of the kind it correlates with.

<sup>5</sup> I present here this overly general characterization of representation so that it will apply to a wide variety of current applications in cognitive science. For a more careful and detailed account see Salay (2019).

<sup>6</sup> Although representation is not the only function of language.

<sup>7</sup> Again, this is overly simplistic since intentional agents often respond to the informational properties of natural phenomena, that is, they treat them as “natural signs.” But this additional observation will not affect the arguments being made here so I will leave discussion of it out.

<sup>8</sup> A great deal has been said about the different ways in which things can function as representations. William Ramsey’s important book, *Representation Reconsidered* (Ramsey, 2007), which neatly distinguishes between the different representational appeals standardly made in cognitive science, is a good place to start. I will make some (implicit) use of his distinction between structural, receptor, and IO-representations here, but, for reasons of space, I will not be making these distinctions an explicit part of this discussion.

But this is a mistake. That association networks are compositional in this way — that the triggering of a part of one will have a cascading effect — is an information by-product of *all* networks. Indeed, it is a by-product of all causal chain processes.

Imagine a Rube Goldberg machine that is initiated when a ball rolls over a lever (X). Depression of X causes a switch (Y) to flip which causes a ramp (Z) to drop, which frees the ball to roll and so on. An information by-product of the mechanical associations of this machine's parts is that the activation of one part is an indication that a certain situation holds. Depression of X, for example, indicates the presence of an X-sized object. But it would be incorrect and not a little misleading to describe X as having the functional role of representing the presence of X-sized objects. Nothing at all in this machine, not even the functioning of the entire machine itself, requires or responds to the information *an X-sized object is present*. The functional role of X here is *to flip switch Y*. That such a machine might be put to further use by some outside agency, say to track the presence of a ball on X, does not change the functional role *for the machine* of the machine's parts. Likewise in the case of classifiers. Although subpatterns of network activity serve to trigger larger patterns of network activity, it would be misleading to describe any of these as functioning to represent aspects of the output class. As in the mechanical system of the Rube Goldberg machine, each part of the response pattern functions to trigger the parts with which it is linked: there is no representation of anything here.

Sentient beings, in contrast, are organized at multiple levels of granularity: lower-level cellular activity yields middle level processes that yield personal level experience. When a cat perceives a mouse, its sensory receptors become active and these in turn trigger a series of neural, respiratory, and muscular responses. At the personal level, the cat becomes aware of the mouse: it sees, smells, and hears it. This awareness, on the standard view, is an internal, mental representation of the mouse that results from the neural processing of the lower-level sensory data. The question here is this: is this mental awareness playing a representational functional role for sentient systems?

If we say yes, as RRH does, we must explain how mental percepts themselves are causes of behavioral responses. But if we analyse mental percepts reductively and take them to be *aspects* of lower-level processes, then sentient perception is no different from non-sentient machine learning: all the causal work is accomplished by the low-level mechanisms. The system is successful in responding to its environment because it has developed responses, at multiple levels of granularity, that reflect the statistical regularities of the kinds in its environment. From a teleological perspective we might find it useful to describe the correlations between the low-level mechanisms and the kinds in the environment as representational, but nothing in such systems is functioning as representations of those kinds for the system itself. On such a reading, then, awareness plays no representational functional role; if it is a *something* at all, it seems to be epiphenomenal. Indeed, the very idea that the reasons that guide intentional behavior can be found in an agent's internal mechanisms rests on a confused picture of the origination of agency. If agent S does something for reason R, then R is (one of) the efficient cause(s) of S's subsequent actions. If perceptions and thoughts function as causes of behavior by way of lower-level mechanisms that encode the information content of this higher level, then it is the mechanism itself that plays the causal role in this picture. For example, if some neural activity, N, functions as a physical implementation of some information, say that there is a fly over there, then N is the cause of S's subsequent actions. But this is a sloppy analysis of how parts contribute to system behavior. Parts of systems do not *cause* system wholes to do anything; they are constitutive of system behavior (Craver & Bechtel, 2007). My feet moving forward are not the cause of my forward movement; they are part of what it is for me to move. My lungs contracting and expanding do not cause me to breathe; they are constitutive of my breathing. Likewise, neural mechanisms are constituents of system behavior, not causes of it.

The crux of the problem with these reductive accounts of representation, I suggest, is the unwarranted coupling of sentient awareness with mental representation. Recall Fred and Sally. For Fred's utterance to function as a representation of *<fire>*, Sally must treat it as such; if Sally simply stared blankly at Fred when he spoke, his exclamation would fail in this regard. Individual behavior becomes representational behavior because of the features of the broader context within which it plays a specific role, that of standing-in for the situation it represents. This larger context lies outside of individual users

and thus no amount of reductive analysis to the lower-level mechanisms of those users could ever comprehensively explain it. While this point is well understood with respect to language — hypotheses about language development typically include accounts of how linguistic practices develop within social communities — the analogous observation about mental representations is generally overlooked.

As Povinelli and Henley observe, phenomenally accessible resources — mental constructs — can and often do play a representational role in the cognitive activities of perceivers. One way with which we are all familiar is the simple act of imagining. I can mentally visualize a red apple on the table even though there is nothing on the table before me. But this kind of cognitive activity is complex, the result of a great deal of learning and, I shall argue, much symbol use experience. The assumption being challenged is that phenomenal awareness is itself a mental representation. Unacknowledged here is the scaffolding required for some bit of awareness to become a *something* at all and further to become a something that functions as a representation. As in the case of language use, establishing a role for, say, perceptions (more precisely, *descriptions* of perceptions) is a coordination problem: an individual must learn to coordinate her own experiences over time. For Sally’s perceptual experience of an apple to function as a representation of <apple> or even <this apple>, Sally must group together her various, temporally distinct, apple experiences. But, as Ludwig Wittgenstein aptly observes (2009, sections 243–279), this is not something Sally, or anyone, can do without the aid of external resources. Once an apple-perceiving experience has passed, it is no longer a *thing* to which a future experience can be compared. To accomplish the comparison, Sally needs an external anchor to which she can attach her various apple experiences. Without this personal level association, the fleeting mental wisps that accompany her apple perceptions are epiphenomenal, at least with respect to a representational function. Sally’s apple experiences emerge out of lower-level response patterns, to be sure, but they take on a representational role *for Sally* only in the context of the larger system within which her experiences can be associated with concept anchors, e.g., words, in concert with the ongoing activity of the symbol using community. This larger system, of course, takes us well beyond the blood-brain barrier of individual perceiving agents.

Lower-level mechanisms thus implement the representation-use behavior of sentient systems, but they cannot do double duty and also explain their representational nature. Once we decouple the key features of the phenomena to be explained, the theoretical approach becomes clear: the problem of sentience requires an internalist, reductive account; the problem of representation-use requires an externalist one.

From this vantage point, we can see that RRH is founded on the same level confusion with which Povinelli and Henley (2020) charge their colleagues, between algorithmic level speculations about internal representational mechanisms and functional level descriptions of representation-use behavior. Thus, contrary to Povinelli and Henley’s claim that “... by necessity, higher-order, structural, role-based representations do entail the presence of first-order, perceptually-based representations” (p. 394), not only do they not, but the appeal to them does not explain the phenomenon at all. As they argue, although many experiments support the *observation* that some animals manifestly wield higher-order capacities, that conclusion cannot be drawn until we rule out the possibility that a combination of first-order operations, perhaps in combination with a fortuitous experiential setup, led to the behavior: “if first-order relational reasoning is both necessary and sufficient [to produce the behavior], then higher-order relational reasoning has no role to play in explaining the results” (p. 406). The human case is no different. Until we can untangle the cognitive advantages afforded by the human cognitive niche – most glaringly language – from biologically fundamental human cognitive capacities, we cannot identify the capacities that humans “manifestly wield.” And without a benchmark set of capacities, comparative work will have no ground on which to stand.

Not surprisingly, much of the support in favor of RRH comes from the developmental literature: “the propensity to evaluate the similarity between states of affairs based on the causal-logical and structural characteristics of the underlying relations rather than on their shared perceptual features appears quite early and spontaneously in all normal humans – as early as 2-5 years of age, depending on the domain and complexity of the task (Gentner, 1977; Goswami, 2001; Halford, 1993; Holyoak et al., 1984; Namy & Gentner, 2002; Rattermann & Gentner, 1998a; Richland et al., 2006)” (Penn et al., 2008a, p.111). But, as the subjects of these studies are all modern human children who have developed within a rich and broad

cognitive niche, it is hard to see what “spontaneous” could mean in this context: is the implication that early life exposure to representation-use, i.e., actively through ostensive learning and passively through observation, is not a prerequisite of this behavior? This is not to suggest that developmental work does not yield important insights: it most certainly does. But it is a reminder that empirical evidence cannot confirm a theory any more than it can determine one. Such case studies support externalism too once we notice that the relevant cognitive behavior emerges at an age when children have been exposed to enough language for linguistic proficiency to begin to develop. On that view, as I shall elaborate shortly, language functions as a representation *tool* and it is skilled use of this tool that ultimately explains the second-order behavior.

But although there is a growing literature of comparative experimental work that supports the externalist view, Penn et al.’s (2008a) treatment of it is cursory and dismissive and Povinelli and Henley (2020) do not mention it at all. In Penn et al. (2008a), externalism is narrowly treated as the idea that language explains the cognitive discontinuity between humans and other animals. They identify and critique three versions of this view: 1) that verbalized (or imaged) natural language sentences are responsible for the disparity between human and nonhuman cognition; (2) that some aspect of our internal “language faculty” is responsible for the disparity; and (3) that the communicative and/or cognitive function of language served as the prime mover in the evolution of the uniquely human features of the human mind (p. 121).

In what follows, I evaluate their critique of 1, the important hypothesis that language extends the cognitive capacities of individuals in specific ways and pass over their discussion of 2 and 3 as these are both versions of internalism.

### **Evaluating the Penn-Holyoak-Povinelli Critique of Externalism**

While the cognitive enhancing capacities of language are acknowledged, “... natural language clearly subserves and catalyzes normal human cognition ...” (Penn et al., 2008a, p. 121), externalism is dismissed on the evidence of one(!) brain-damaged patient, “an agrammatic aphasic man who was incapable of producing or comprehending sentences and whose vocabulary was essentially limited to perceptual nouns” (p. 121),<sup>9</sup> the evidence of congenitally deaf children who “spontaneously” develop “gestural languages with hierarchical and compositional structure” (p. 121), and what they describe as the failed attempts to teach language to nonhuman animals.

With respect to the limited studies of deaf children, what force could the claim of spontaneity have once we recognise that each of these children has grown up in an environment that is replete with language — in the digestible visual form of written words and gestures — and linguistic cues — again, visual, in the form of implicit and explicit invitations to share eye gazes? Deaf children do not develop outside of the human linguistic cognitive niche. Nevertheless, the lesson we are to take away from this “case” is “confirmation that the human mind is indomitably human even in the absence of normal linguistic enculturation” (Penn et al., 2008a).

As much as the evidence in favor is overblown, the evidence to the contrary is given short shrift with a quick and sweeping indictment of even the possibility of animal language use justified by five ‘failed’ animal language projects. Except for (Pepperberg, 2002), none of these is from the current century. It is true that human language development is glaringly conspicuous in comparison with that of nonhuman animals. And it is indeed extremely likely that this contrast is partly due to cognitive capacities that set humans apart. Just what these are, however, and the extent to which language or simple symbol-use is possible for nonhuman animals are questions we can answer conclusively only with more research. And this research is currently ongoing (Arnold & Zuberbühler, 2008; Beecher, 2021; Bergman et al., 2019; Clay & Zuberbühler, 2009; Fedurek et al., 2016; Fitch, 2020; Grüter & Czaczkas, 2019; Herman, 2006; Herman et al., 1984; Pratt, 2019; Searcy & Nowicki, 2019; Ten Cate, 2017; Ten Cate & Spierings, 2019). Indeed,

---

<sup>9</sup> The study is described in Varley and Siegal (2000).

as the theoretical paradigm shifts toward externalism, observations that were not possible before, because the internalist lens blinded us to them, could reveal linguistic animal behavior where we once saw none.<sup>10</sup>

Finally, the way in which studies such as Thompson and Oden (2000) and Thompson et al. (1997) support the externalist claim are misconstrued in the critique. They are not meant to provide "... evidence that symbol-trained animals are ... more adept than symbol-naïve ones at reasoning about unobservable causal forces, mental states, analogical inferences, or any of the other tasks that require the ability to cognize higher-order relations in a systematic, structural fashion" (Penn et al., 2008a, p. 122). The externalist claim is that representational tools such as symbols reduce second-order problems to first-order ones. Higher-order problem solving is made possible by such simplifications of the problem space, for humans and nonhuman animals alike. If this is right, then we expect both symbol-trained subjects as well as symbol-naïve ones who are trained to use problem-relevant symbols to perform about the same on the problem-solving tasks. This is precisely what the studies show: once the symbols are learned, the higher-order cognitive behavior emerges. By viewing the studies through an internalist lens, however, Penn et al. (2008a) see them as failures because they do not support the conclusion that there is an internal difference between symbol-trained and symbol-naïve subjects, one that will allow the former, but not the latter, to succeed at tasks that, they assume, "... require the ability to cognize higher-order relations in a systematic, structural fashion" (p. 122).

Having found no reason against the viability of externalism, I turn now to a version of it that offers an alternative analysis of our second-order cognitive behavior, one that directly challenges the relational hypothesis of RRH.

### Externalism

The guiding principle of the externalist approach is parsimony: assume only what is required to make sense of behavior. Thus, while the starting point is an acknowledgment that representation has a critical role to play in higher level cognition, externalism does not assume that there must be an internal mechanism playing that role; indeed, as we have seen, it begins with a rejection of the standard internalist treatment of basic perceptual awareness as representational. On externalism, in contrast, perception is analysed as a relational process rather than as an informational one.

### Relational Perception

According to relational theories of perception (Brewer, 2017; Hurley, 1998; Martin, 2002; Noë, 2010; O'Regan & Noë, 2001; Travis, 2013), perception is an interactive process, something that animals *do*. With locomotion, animals move through their environments; with perception, animals engage with salient aspects of it. Saliency is partly determined by the world and partly by the embodiment details and sensorimotor contingencies of an animal's sensory capacities. During vision, for example, the sensory stimulation of retinal photoreceptors will shift and change in lawful ways as eyes rotate. If eyes are located on the front of the body, the "flow of information" will expand and contract as the animal moves forward and backward. The idiosyncrasies of an animal's sensory apparatus in conjunction with the environment in which it develops dictates its perceptual expectations which, in turn, drive the animal to adjust its body so that it is optimally oriented toward the objects with which it is engaging. Perception is the ongoing, *skilled*, exercising of these sensory modes of environmental engagement, the dynamic interchange between a perceiver and its world.

When we let go of the idea that perception is the internal construction of representations that a perceiver "has," we are no longer tempted to treat low-level activity as pseudo representational either. On this view, sensory activity is a combination of instinctive and learned responses to environmental stimuli that together constitute an animal's capacity to interact with its environment. We tend to use pictorial

---

<sup>10</sup> My purpose here is not to defend against those who are skeptical of the possibility of nonhuman language use, but to point out that the debate is ongoing and that an out-of-hand dismissal of the possibility is not a tenable position.



language when describing perceptual experiences – “I have a red, round image” – but it is only the descriptions that are representations. When perceiving, animals are physically interacting with real objects that are whole and continuously present: “what explains the conceptual unity of experience is the fact that experience is a thing we are doing, and we are doing it with respect to a conceptually unified external object” (O'Regan & Noë, 2001, p. 967). Thus, simple perception of its world can cause an animal to act, but as perception is an interaction on this view, it is the object of engagement, a thing in the world, that is the causal progenitor of that action.

But what of obviously mental acts of representation manipulation<sup>11</sup> such as when Sally mentally rotates the figure displayed on the paper to determine the correct next choice in the series? This and others like it are indeed clear examples of mental representation-use, but as I noted earlier, it does not follow from the fact that we can experience our own representation-use behavior (because we are sentient beings) that the function of mental awareness is to represent. Without further scaffolding, the wispy memory traces of perceptual experiences are just one aspect of a sentient being's response pattern, across levels of granularity, to some perceptual interaction. A capacity for unified experience is a by-product of sentience, but it does not entail a capacity to *use* these experiences. In other words, sentience is a necessary but not a sufficient condition of mental representation-use. We still need an account of how this use develops. On the externalist analysis, such a capacity is grounded in the use of physical representations in the world, initially natural signs and then, as complexity increases, increasingly arbitrary symbols. To understand how the internal capacity develops, then, we need to first understand how representation tool-use develops. Elsewhere (Salay, 2019), I give a detailed account of the critical, non-intentional, prerequisites for becoming a sign-tool-user. After a sketch of that work, I explain how this analysis offers a novel answer to the problem of second-order relational reasoning and opens new avenues for comparative research that are not prey to the Povinelli-Henley critique.

## Representation Tools

Animals learn through experience with their environments, either over an evolutionary timescale across a species, or over a developmental timescale across an individual's life. To understand how *representation* tool-use emerges, then, we should look for both phylogenetic and ontogenetic factors that support and prompt its development. Much has been written on this topic and here I highlight only some salient central ideas.

An obvious precondition of the capacity to learn and use sign tools — to “share intentionality” (Tomasello et al., 2005) — is the ability to learn to use tools more generally.<sup>12</sup> At a minimum, it also seems to require phylogenetic traits such as sociability, an inclination to follow eye gaze, and a capacity for interpreting the perceptions of others (Baron-Cohen, 1995; Call et al., 2005; Csibra & Gergely, 2007; Gómez, 2009; Savage-Rumbaugh, 1990; Wilson & Sperber, 2002). These sociability traits alone, however, are not enough. As Tomasello et al. (2005) point out, the fact that an animal has the cognitive capacity to recognise and understand the intentions of others is “not by itself sufficient to produce humanlike social and cultural activities” (p. 676). There are many instances of intention recognition across species; nevertheless, very few have developed the sort of intentionally rich social world that humans have. Without an established social *practice* of intention-sharing, individuals will not develop intentions to share. In other words, the relationship between developing individuals and their sociocognitive niche forms a virtuous

---

<sup>11</sup> Note that I am not speaking here of the capacity to correctly identify perceptual changes in the environment, as in change-blindness studies (e.g., Hollingworth & Henderson, 2002; Mitroff et al., 2002). When these are the result of wholly unconscious processes, the inner mechanisms are not playing a representational role for the system although it is useful for us, third person observers, to describe the relationship between the inner mechanisms and the outer environment as a representational one. The situation is complicated in the human case since we have language and thus, a human subject's report of memory will be a linguistic response. This, as we shall see shortly, is precisely where representations enter the picture, not from the lower-level mechanisms, but from the external, linguistic infrastructure that pervades the human's world.

<sup>12</sup> There is a rich literature on tool use across species. For an excellent review, see Bentley-Condit and Smith (2010).

circle: there is no niche without individuals who engage in intention-sharing behavior, but intentions themselves emerge out of these ongoing interactions.

It is very unlikely, in our view, that a human or ape kept in social isolation for the first year of life would suddenly understand others as goal-directed or intentional agents on its initial encounter with them; presumably the developmental pathway for understanding intentional action depends on species-typical social interactions early in ontogeny (Tomasello et al., 2005, p. 688).

Thus, an innate propensity to engage with members of a social group *in the context* of an intentionally rich social world is the ground of intentional group dynamics, an ongoing reciprocal interaction that creates situations "... which make some aspects of entities in the shared situation 'mutually manifest' and so potentially 'relevant' for acts of interpersonal communication (Sperber & Wilson, 1986)" (cited in Tomasello et al., 2005, p. 683). Intentions emerge out of the learned practice of behaving in certain ways in certain situations. Without this social ground, individual behavior *means* nothing at all.

In the context of such an intentional social niche, behavior that exploits natural signs – the regular cause and effect relations in the environment (Grice, 1989) – can become a potent source of information for the entire group. Smoke, for example, is a physical effect of fire and, thereby, a natural sign of it. An animal that is capable of perceiving smoke could learn, given the right sort of conditioning experience, to respond to it in the way it does to fire, presumably with avoidance. Seeing a plume of smoke rising from the forest canopy, smelling a change in the atmosphere, the wolf moves to high alert, fur bristling, ears pricked, muscles flexed, ready to flee. Immediately, the rest of the pack responds in kind, searching for the source of the danger. Because the wolf has learned to expect fire when perceiving smoke and because it is part of a social world in which sharing intentionality is possible, smoke takes on a representational role for the group: it is a sign of fire.

Visual and olfactory changes during estrus, phenotype changes in the face of threats, and instinctive vocalisations in predatory situations are all examples of natural signs to which many animals instinctively respond. As such, they play rudimentary representational roles in the relevant contexts. But representation becomes an increasingly significant aspect of social situations with the development of simple sign *tools* – local, ontogenetic ritualizations that must be learned through social transmission: "infant-mother dyads, for example, will develop idiosyncratic, stylised carry signals, e.g., a shoulder touch, that trigger responses, carrying behavior, that would normally be triggered by overt carry request behavior, e.g., climbing onto mother's back" (Salay, 2019, pp. 137-138). More complex sign tools are those that are highly manufactured, namely arbitrary signs whose physical properties have nothing in common with their representational properties. The use of the sequence of sounds /bɔ:l/ to indicate instances of <ball> is one example.

In the context of a social animal, a capacity to use signs, even natural ones, yields the possibility of shared information and then of the development of shared information practices. Once a system of such practices is developed, the dynamic interchange between sign tools and the environmental pressure to produce them fuels an expansion of both tool development and the practice of them. As signs become increasingly manufactured, the constraints around sign use possibilities become less tied to the features of the physical environment and more to the social one. When such signs are easy to produce, the result can be an explosive proliferation of them.

For humans, vocalisation is a natural medium of representation. As some have hypothesized (Lieberman, 1984, 2002; Negus, 1949), the early development of rudimentary vocal signs might have spurred the evolution of further vocal structural development, which in turn, supported more nuanced vocalisations. The landscape that humans develop in is thus one that is replete with words. For a child, developing in the context of a cognitive niche in which there is a widespread, social practice of language use, learning how to use these tools is a critical precondition for interaction with the members of her social world. Indeed, there are more words in a child's environment than there is anything else. From a phenomenological perspective, words are "ready-to-hand" (Heidegger, 1962) tools that draw users in the same way that, "if it gets too warm, the windows solicit us to open them" (Dreyfus, 2007, p. 1158). Seeing the ball by the swings, Sally picks it up and begins to throw it, yelling "Ball!" as she does so, thereby attracting other children to join in. In this spontaneous and interactive way, the ball-playing unfolds. According to externalism, there is not first an inner representation of some *desire to play ball* that Sally

conveys with her utterance. Rather, the presence of the ball (in the context of play, in a niche in which ball play is a practice, together with the ready availability of the word “ball”) provokes Sally to shout “Ball!”.

As an individual’s experience with words increases, many and varied word associations develop as well. A conductor who routinely lifts her right hand and arm in a particular way at the opening of a piece of music might experience rapid, episodic flashes of past conducting experiences as she reaches toward the top shelf of her bookcase. Her own body movement has become associated with *<conducting situation>* and now itself triggers her multi-level response to instances of them. Likewise, using or hearing a familiar word or phrase can trigger episodic memories and, because neural activity cascades along networks of synaptic associations, all associated subpatterns will become active to some degree as well. Some of these will involve other word-use experiences thus leading to the steady stream of mental activity that we experience when we stop to pay attention to it.

To sum up, a context in which there is a social practice of using representation tools is a necessary precondition of mental representation: without this external scaffolding, the wispy mental images that seem to make up our minds are merely traces of past perceptions, that is, real-time, ongoing, interactions between agent and world. To think is to mentally use words according to the representational role they play in the wider linguistic community. Words are “out there” in the possibilities of their use, but they seem to be “in here” because they are so easy to produce and there are so many of them roiling about in our consciousnesses. Sentience is thus a necessary but not sufficient ground of mental representation: it develops only in the context of a cognitive niche in which there is a practice of shared intentionality and the development of representation tools. As sign use develops, these tools become increasingly ready-to-hand and thereby a draw for further use. The degree to which an animal can use representations will depend in part on how widespread the social practice of representation tool-use is in its cognitive niche. I turn now to a discussion of how representation tools enable second-order relational reasoning.

## Second-Order Relational Reasoning

The first move toward an externalist account of second-order judgements is to drop the assumption that they require second-order capacities. As extended mind theorists are fond of pointing out (Cimatti & Vallortigara, 2015; Clark, 2006a, 2006b; Deacon, 1998; Donald, 1991; Logan, 2007; Wheeler, 2004), simplifying problems that beings with brains such as ours are not suited to solving is one of the key cognitive-enhancing aspects of the “widewear” that helps us navigate our physical and social environments: “Experience with external tags and labels thus enables the brain itself – by shallowly representing those tags and labels – to solve problems whose level of complexity and abstraction would otherwise leave us baffled” (Clark, 2006b, p. 294). Most human beings are incapable of complex mental calculations, but by representing numbers with symbols and displaying them as in Figure 1, even young children can solve difficult mathematical problems. We use representations in this simplifying way each time we put pen to paper, use external cues as reminders, keep ourselves on task with self-talk, and myriad other mundane daily actions we perform without our even being aware of them.<sup>13</sup> According to externalism, then, we can solve second-order problems *because we have tools that reduce them to first-order ones*.

**Figure 1**

*Long Multiplication Technique*

$$\begin{array}{r} 8963 \\ \times 465 \\ \hline 44815 \\ 537780 \\ 3585200 \\ 4167795 \end{array}$$

<sup>13</sup> Simplifying problems is not the only way representations enhance our cognitive capacities: they enable tracking and thereby responsibility to “offline” features of our environment; they make unemotional decision-making possible; they make abstract thought possible. More on this last presently.

Sally is trying to solve a difficult logical problem but there are too many variables for her to track. She gets out a notepad and, using symbols for salient aspects of the problem, she develops a representation, as illustrated in Figure 2, of the problem space. Here her model is functioning as a representation of the problem, as a stand-in for the original problem. Because some of the details of the original are stripped away, she can see relationships that before were opaque and she solves the problem easily. For simple problems, Sally might make do with a mental problem model. Whether the model is displayed on paper or in her imagination, it is functioning as a representation of the problem itself.

**Figure 2**

*Logical Form Notation*

$$(A \& B) \vee (C \& D)$$

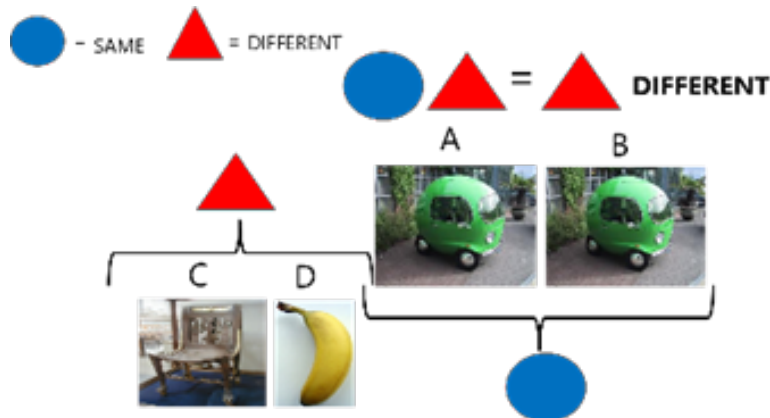
$$\sim(A \& B)$$

$$\therefore (C \& D)$$

As Clark (2006b, p. 294) observes, even simple symbol tokens can convert difficult second-order tasks into simple first-order ones. Language-naïve chimpanzees that are taught to associate relations such as <same> and <different> with symbol tokens are subsequently capable of solving the hitherto opaque second-order task of identifying whether *groupings* are of the same or different kind. (Thompson et al., 1997). See Figure 3 for an example. These chimpanzees have not thereby gained a new capacity, an inner understanding of some abstract relation; rather, with the aid of the symbol tokens the second-order problem is reduced to the simple first-order one of comparing tokens.

**Figure 3**

*Plastic Tokens Associated with Pairs of Pictures that are the Same and Different, Respectively*



Language does not endow us with a second-order capacity either; it reduces second-order tasks to first-order ones. Words are stand-ins for complex, abstract, spatio-temporally extended norms of use. We are not the sorts of beings that can interact with such entities, but by developing responses to the words that stand for them, we are able to solve problems that require them.

On RRH, our ability to mentally manipulate “objects” – basic representations that are “given” in perception – is an internal capacity that humans are born with, not something that requires a community-wide practice of representation-use that individuals must learn. Sally’s perception of a cat on a mat is constituted by a neuro/mental construction of cat and mat representations. In *judging* that the cat representation is related to the mat representation via the <on> relation, she is invoking her second-order neuro/mental processing capacity. A “gap” between the low-level neural processes that implement the computations that yield the Sally-level judgements is a consequence of such reductive analyses. The intractable problem of giving an account of it has been labelled “the hard problem” (Chalmers, 1995), a

problem so hard that most researchers ignore it. On the externalist view, in contrast, because it does not treat Sally's low-level processes as the causal progenitors of her mental representations, there is no gap to be crossed: Sally is a whole animal that is constituted by myriad processes at multiple levels of granularity. When Sally perceives the cat and the mat, her "perception" of the cat and the mat include the actual cat and the mat: perception is an interaction, not an internal, mental activity. How does she judge that the cat is <on> the mat? On externalism, the ability to recognise relations between the objects we engage with in perception is not a natural human capacity: it is a consequence of a great deal of learning experience in the context of a representationally rich cognitive niche.

Consider, for example, the simple relation <living in>. We expect young children to be capable of making judgements of the following sort: humans live in houses; birds live in nests; bears live in caves. How do they learn these? The answer for each child will be different, of course, since each has a unique learning history. But we can imagine a young Sally leafing through her picture books and seeing pictures of birds in nests next to picture of bears in caves and people in homes. Sally may not have seen an actual bear in a cave and a bird in a nest, but she has surely had experience of living in her own home and, as she looks at the various pictures, she responds to them in the ways that she has learned. Because sitting and looking at the pictures in her book, perhaps with her mother at bedtime, is itself also a situation that Sally is engaged in, she is developing new associations with those already entrenched ones: her current response to the pictures of birds in nests, bears in caves, and people in homes is now mixed up with her previous responses to homes. Sally does not have an inner <live in> concept, a model of what it means to stand in the relation of <living in> to a structure; rather, Sally has learned a new response to <living in> situations, one that is triggered not just by the sight of her own home, but by birds in nests and bears in caves as well. Eventually, Sally will learn to associate a label with these situations so that, in addition to pictures of birds in nests, bears in caves, and people in homes, strings of letters and sounds — "live in" — will also function to trigger these responses. With a label for it, something that turns a second-order relation into a concrete thing to react to, she will be able to think *about* the relation itself, just as chimpanzees (Thompson et al., 1997) were able to identify second-order relations with the introduction of symbols. Representation tools make talking and thinking *about* the stuff of our world possible. Sally herself has not gained a new second-order capacity; rather, it is Sally + her representation tools that has the second-order capacity.

Coming back to the original problem facing comparative researchers, what can externalism teach us about how we might identify higher-order cognition in nonhuman animals? Povinelli and Henley argue that no conclusions can be drawn if and until the possibility of learned, first-order responses to the experimental situation has been ruled out. On externalism, however, there are only first-order responses: no animal, not even humans, has biologically endowed, second-order cognitive capacities. Those animals that can solve second-order problems do so with the aid of external scaffolding that reduces the complexity of problems to first-order ones.

## Conclusion

The externalist answer to the Povinelli-Henley challenge, then, can be described in Marr's terms as follows: Functional Description – humans make use of second-order concepts to solve complex problems; Algorithmic Specification of Supporting Mechanism – complex problems are simplified into first-order problems with the aid of representation tools, namely words and sentences, that function as placeholders for second-order relations.<sup>14</sup> The testable hypothesis that follows from externalism is this: an animal's higher-order cognitive capacity is a function of both its capacity to learn to use representation tools and the complexity of the tools at its disposal. Directly testing for (or observing) ability to use representation tools at all, then, must be the first step in any externalist comparative study. Next steps would investigate whether complexity of cognitive behavior correlates with increase in number and

---

<sup>14</sup> For a detailed account of how these representation tools are developed and learned without some a priori capacity for apprehending them see Salay (2019). The general idea is that increasingly complex representations become possible once a groundwork of simple representations is established. In all cases, understanding is cashed out as proficiency in use.

complexity of representation tools available. If externalism is right, humans do not have cognitive capacities that are magnitudes greater than those of nonhuman animals; rather, their advanced cognitive behavior emerges out of their cognitive niche, rich in representational tools.

**Conflict of interest:** The author declares no conflict of interest.

### Acknowledgements

Thank you to valuable feedback from Jennifer Vonk and Ben Farrar: responding to their suggestions has made this a much better paper. Thank you as well to those who produce and contribute to this most excellent journal. These articles have played a critical role in directing my current thinking about cognition.

### References

- Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), 202–203.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.
- Beecher, M. (2021). Why are no animal communication systems simple languages? *Frontiers in Psychology*, *12*.
- Bentley-Condit, V., & Smith, E. (2010). Animal tool use: current definitions and an updated comprehensive catalog. *Behaviour*, *147*(2), 185–227.
- Bergman, T., Beehner, J., Painter, M., & Gustison, M. (2019). The speech-like properties of nonhuman primate vocalizations. *Animal Behaviour*, *151*, 229–237.
- Berwick, R., & Chomsky, N. (2015). *Why only us: Language and evolution*. MIT Press.
- Brewer, B. (2017). The object view of perception. *Topoi*, *36*, 215–227.
- Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Animal Cognition*, *8*(3), 151–63.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.
- Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*, 1387–1396.
- Cimatti, F., & Vallortigara, G. (2015). So little brain, so much mind. Intelligence and behaviour in non human animals. *Reti, saperi, linguaggi*, *2*(1), 5–22.
- Clark, A. (2006a). Language, embodiment, and the cognitive niche. *Trends in Cognitive Science*, *10*(8), 370–374.
- Clark, A. (2006b). Material symbols. *Philosophical Psychology*, *19*(3), 291–307.
- Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, *101*, 401–431.
- Craver, C., & Bechtel, W. (2007). Top–down causation without top–down causes. *Biology and Philosophy*, *22*, 547–563.
- Csibra, G., & Gergely, G. (2007). *Ember és kultúra: A kulturális tudás eredete és átadásának mechanizmusai*, *11*. Akadémiai Kiadó.
- Deacon, T. (1998). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Company.
- Donald, M. (1991). *Origins of the Modern Mind*. Harvard University Press.
- Dreyfus, H. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence*, *171*, 1137–1160.
- Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, *6*(38226), 1–11.
- Fitch, T. (2020). Animal cognition and the evolution of human language: why we cannot focus solely on communication. *Philosophical Transactions of the Royal Society*, *B375*, 20190046.
- Fragaszy, D., & Mangalam, M. (2020). Folk physics in the twenty-first century: Understanding tooling as embodied. *Animal Behavior and Cognition*, *7*(3), 457–473.
- Freeman, W. (2000). *How brains make up their minds*. Columbia University Press.
- Gómez, J. C. (2009). Embodying meaning: Insights from primates, autism, and Brentano. *Neural Networks*, *22*(2), 190–196.
- Grice, H. (1989). *Studies in the way of words*. Harvard University Press.
- Grüter, C., & Czaczkes, T. J. (2019). Communication in social insects and how it is shaped by individual experience. *Animal Behaviour*, *151*, 207–215.

- Heidegger, M. (1962). *Being and Time* (J. Macquarrie & E. Robinson, Trans.). Harper & Row. (originally published 1927)
- Henley, T. (2020). Do New Caledonian crows represent <weight>? An Analysis of Jelbert et al. (2019). *Animal Behavior and Cognition*, 7(4), 661–664.
- Herman, L. (2006). Intelligence and rational behavior in the bottlenosed dolphin. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 439–367). Oxford University Press.
- Herman, L., Richards, D., & Wolz, J. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, 16, 129–219.
- Hollingworth, A., & Henderson, J. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136.
- Hurley, S. (1998). *Consciousness in action*. Harvard University Press.
- Jelbert, S., Miller, R., Schiestl, M., Boeckle, M., Cheke, L., Gray, R., Taylor A., & Clayton, N. (2019). New Caledonian crows infer the weight of objects from observing their movements in a breeze. *Proceedings of the Royal Society B*, 286(1894), 20182332.
- Keijzer, F. (1998). Doing without representations which specify what to do. *Philosophical Psychology*, II(3), 269–302.
- Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology*, 45, 36–62.
- Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.
- Logan, R. (2007). *The Extended Mind: The Emergence of Language, the Human Mind, and Culture*. University of Toronto Press.
- Marr, D. (1982). *Vision*. W. H. Freeman and Company.
- Martin, M. (2002). The transparency of experience. *Mind & Language*, 17(4), 376–425.
- Mitroff, S., Simons, D., & Franconeri, S. (2002). The siren song of implicit change detection. *Journal of Experimental Psychology: Human Perception and Performance*, 28(4), 798–815.
- Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. Hafner.
- Noë, A. (2010). Vision without representation. In N. Gangopadhyay, M. Madary, and F. Spicer (Eds.), *Perception, action, and consciousness: Sensorimotor dynamics and two visual systems* (pp. 245–256). Oxford University Press.
- O’Regan, K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939–1031.
- Penn, D., Holyoak, K., & Povinelli, D. (2008a). Darwin’s mistake: explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31(2), 109–130.
- Penn, D., Holyoak, K., & Povinelli, D. (2008b). Darwin’s triumph: Explaining the uniqueness of the human mind without a deus ex machina. *Behavioral and Brain Sciences*, 31(2), 153–178.
- Pepperberg, I. M. (2002). *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Harvard University Press.
- Povinelli, D., & Henley, T. (2020). More rope tricks reveal why more task variants will never lead to strong inferences about higher-order causal reasoning in chimpanzees. *Animal Behavior and Cognition*, 7(3), 392–418.
- Prat, Y. (2019). Animals have no language, and humans are animals too. *Perspectives on Psychological Science*, 14(5), 885–893.
- Ramsey, W. (2007). *Representation Reconsidered*. Cambridge University Press.
- Salay, N. (2021). How reductive analyses of content are confused and how to fix them: A critique of varitel semantics. *The Journal of Mind and Behavior*, 42(2), 109–138.
- Salay, N. (2019). Learning how to represent: An associationist account. *The Journal of Mind and Behavior*, 40(2), 121–145.
- Savage-Rumbaugh, S. (1990). Language as a cause-effect communication system. *Philosophical Psychology*, 3, 55–76.
- Searcy, W. A., & Nowicki, S. (2019). Birdsong learning, avian cognition and the evolution of language. *Animal Behaviour*, 151, 217–227.
- Ten Cate C. (2017). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin & Review*, 24(1), 91–96.
- Ten Cate, C., & Spierings, M. (2019). Rules, rhythm and grouping: auditory pattern perception by birds. *Animal Behaviour*, 151, 249–257.
- Thompson, R., & Oden, D. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, 24, 363–96.

- Thompson, R., Oden, D., & Boysen, S. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, 23(1), 31–43.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–735.
- Travis, C. (2013). The silence of the senses (revised). In *Perception: Essays after Frege* (pp. 23–58). Oxford University Press.
- Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, 92(7), 345–381.
- Varley, R. A., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and “theory of mind” in an agrammatic aphasic patient. *Current Biology* 10(12), 723–726.
- Wasserman, E. (2008). On possible discontinuities between human and nonhuman minds. *Behavioral and Brain Sciences*, 31(2), 151–152.
- Wheeler, M., (2004). Is language the ultimate artefact? *Language Sciences*, 26, 693–715.
- Wilson, D., & Sperber, D. (2002). Relevance theory. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Blackwell.
- Wittgenstein, L. (2009). *Philosophical investigations*, 4th edition. P.M.S. Hacker and J. Schulte (Eds.). Wiley-Blackwell. (Original work published 1953).



## Appendix

One of the reviewers<sup>15</sup> suggested that an externalist analysis of a recent study exploring higher order abilities would help readers understand the implications of externalism to future work. As Henley (2020) provides an RRH analysis of Jelbert et al. (2019), a study that investigates whether New Caledonian crows represent the second-order concept *<weight>*, an externalist appraisal of Henley's analysis will help to clarify the distinctions between the two approaches.

The subjects — twelve, wild, New Caledonian crows — were trained to lift both light and heavy objects and to drop them down a transparent tube. Subjects were divided into two groups, those that would be rewarded for dropping heavy objects down the tube and those that would be rewarded for dropping light objects. Henley (2020) focuses on the group rewarded for dropping the heavy objects.

In the test phase, the subjects were shown two new and visually distinct objects (they were not able to touch these objects), one heavy and one light, suspended by strings. A fan was used to make the light object swing on its string while the heavier one remained motionless. The objects were then removed from the strings and presented to the subjects. On the first trial, a statistically significant majority of the crows that had been rewarded with dropping heavy objects down the tube chose the heavy object, the one not rotating on the string. When there was no fan, on the other hand, the birds showed no preference for the heavier object. Jelbert et al. (2019) conclude that the crows' behavior was governed by an inference involving the second-order concept *<weight>*: "New Caledonian crows used pre-existing knowledge about the behavior exhibited by differently weighted objects in the wind, to infer whether the novel objects were heavy or light, then used this information to guide their choices" (pp. 5-6).

Henley (2020) disagrees. On his analysis, the behavior demonstrates only that the crows used a first-order perceptual representation of "the relation between *effort-to-lift* and *spontaneous-displacement-of-objects*" (p. 663). Henley's neat reconstruction of the explicit and hidden premises in the argument is worth reproducing in full here.

P1 When presented with multiple two-object sets, with one in every pair requiring high *effort-while-lifting* and one with low *effort-while-lifting*, wild-caught New Caledonian crows learn to reliably choose the object with high *effort-while-lifting*.

**P2 Wild-caught crows have had innumerable experiences wherein objects with varying degrees of spontaneously displacement and *effort-while-lifting*.**

**P3 *Effort-while-lifting* physiologically programs *effort-to-lift*.**<sup>2</sup>

P4 When presented with objects that spontaneously move or do not move, crows trained to pick objects with high *effort-to-lift* select the object that did not move.

C1 The crows possess a perceptual representation of the relation between *effort-to-lift* and *spontaneous-displacement-of-objects*.

C2 The crows possess the higher-order representation, *<weight>*.

*Logical weakness:* The capacities referenced in C1 are necessary and sufficient to produce the results, therefore C2 is unnecessary. (Henley, 2020, pp. 662-663).

Henley's central claim is that the original conclusion, C2, is too strong: once we supply P3, only C1 is warranted. An externalist meta-analysis agrees with this assessment but diverges in its characterization of C1. On externalism, as I have explained, perception is an interactive process: to visually perceive an object is to physically interact with it. Importantly, it is not, as the standard information-processing analysis has it, the construction of an internal *representation* of the object of perception. As with all activity, behavior that has been learned in certain contexts will be primed for in similar future contexts. This learning will include the experimental training that primes subjects for *little-effort-to-lift* when interacting with light objects and *high-effort-to-lift* when interacting with heavy objects. It will also include the host of in-the-wild training, captured by P2, which primes subjects for *little-effort-to-lift* when interacting with spontaneously displaced objects. There is no need, given this analysis, to invoke a perceptual representation posit to explain subjects' preferences for motionless objects. If spontaneous

<sup>15</sup> Thank you to Ben Farrar for this excellent idea and for suggesting these papers.

displacement of objects primes *little-effort-to-lift* and subjects are rewarded in situations involving *high-effort-to-lift*, spontaneously displaced objects will repel them. Likely there is a push and a pull here: the spontaneously displaced objects repel the birds and the stationary ones, possibly because they are the remaining alternative or possibly because there is a corresponding priming of *high-effort-to-lift*, attract them. Externalism thus characterizes the finding like this:

P1-P4 As above.

C1 Experience with spontaneously displaced objects primes *little-effort-to-lift*.