1 Introduction.

Suppose a group of children that, for whatever reason, wants to explain the movements of dolls and puppets. Suppose they take two systems of interest to study – a marionette, and a more new-fangled doll that moves and talks on its own. These behaviors of these two systems will require very different explanations. In the marionette, the motive force for its movements originates *outside* of its boundaries; namely, with the strings and the puppeteer. In the modern doll, however, the explanation of its movements will posit accounts about its *internal* organization – i.e., its motor, its microphone, etc. In the terminology we will employ in this paper, the marionette's movement has an *external* locus of control. The doll's behavior has an *internal* locus of control.

We will argue that long-running debates in evolutionary biology and cognitive science have taken place between *internalists* and *externalists* about the locus of control. While these terms have a long history in philosophy, for instance with regards to semantics and epistemic justification, we only employ them here for thinking about the relative causal contributions of components internal to a system and those external to it to that system's behavior, as described in the case above (i.e., we do not mean to imply anything about those other debates). In this dialectical setting, "internalists" argue that the causal source of the phenomenon of interest is internal to the organism, and "externalists" that it is external to it. In evolutionary biology, one phenomenon of interest is phenotypic evolution. Internalists have historically focused on genetic modification as the causal source of phenotypic change, while externalists focus on the environment and developmental organization of an organism. In cognitive science, the phenomenon of interest is intelligent behavior. Internalists have generally focused on explaining behavior by positing internal representations or models, while externalists deny the importance of these constructs, instead positing interaction with the environment as the primary *explanans*.

Interestingly, in the evolutionary biology debates there is an emerging consensus towards what we call a *distributed* locus of control. Modern theorists have begun to recognize that both genes and environmental factors exert specific, but non-redundant, causal influence on the heritable variations. No such consensus has developed in the cognitive science case, however. We suggest that such a move is needed. We further suggest that moving to a distributed locus of control has upshot for the construct of 'mental representation'. In evolutionary biology, the concept of 'gene' has been modified with the progress of genetic and molecular investigation. Although a plurality of gene concepts can be recognized today in evolutionary biology, they are not anymore seen as a causally sufficient internal program for producing traits. In contrast, a view based on a distributed locus of control interpret genes as a non-sufficient causal contributor that interacts reciprocally with environmental factors (Waters, 2006; Noble, 2008). Again, we suggest that a similar concept change can productively posit mental representation as the internal contributor to a distributed locus of control for behavior.

At the outset, we note that positing a distributed locus of control does not end debate. In any specific case, it is still important to determine the relative causal contributions of internal and external factors for the phenomenon of interest. Moreover, individual scientists may reasonably focus, both methodologically and theoretically, on primarily one kind of causal factor. What it does do, and what we suggest has occurred in evolutionary biology, is produce a *minimal agreement*, that in many or most cases the causal source of the phenomenon will involve, non-eliminably, both internal and external factors. On a distributed-locus view, it is thus *methodologically* important to decompose systems into internal and external factors, study their interactions, and their relative contribution in different contexts.

We begin (2) by defining a locus of control, and explaining the difference between *internal*, *external*, and *distributed* loci. In section (3), we assess the history of debates in evolutionary biology, arguing that the Modern Synthesis primarily posited an internal locus of control for genetic variation (3.1), that critics of the modern synthesis proposed alternatives based on external loci (3.2), and that contemporary evolutionary theory is moving towards a minimal agreement that the locus of control for evolutionary change is distributed. In section (4) we introduce the debate surrounding "Enactivism" in cognitive science, and argue that, here, the debate is similarly driven by different sides positing different loci of control. In section (5), we propose how a distributed-locus view accounts for behavior and describe the role of mental representation in these explanations, giving examples from psychology and neuroscience. Section (6) concludes.

2. Locus of Control

A locus of control is a combination of *causal* and *locational* factors. It is causal in the sense that manipulating the locus of control changes the phenomenon of interest in a specific, fine-grained way. It is locational because it occupies some position, either spatial or topological, in the system under study. Suppose that the phenomenon of interest is the long-term fiscal strategy of a corporation. In this case, the locus of control is the board and the CFO. Variations in the decisions made by these factors will change the fiscal policy in a fine-grained way, and these factors are located at a specific location in the organization, namely at the top of a decision hierarchy.

We can further flesh out the notion of causal control of the phenomenon with the notion of *causal specificity*, familiar from the literature on causation (Woodward, 2010; cf. Klein, 2017). According to interventionist accounts of causation, causes are specific to effects when fine-grained manipulations of the cause result in fine-grained changes to the effect. Woodward's example is a radio: while the power switch will exert coarse control over the noise emitted from the radio – either sound or no sound – the tuning dial will exert fine-grained control by determining the specific frequency to which the radio is tuned. On our view, a locus of control comprises components that exhibit fine-grained causal influence on the phenomenon of interest.

The notion of 'location' we care about in this paper is whether the locus of control is *internal* or *external*. Internalists localize the locus of control within an organism, externalists outside of it. This needs immediate further specification, though, since organism-environment boundaries are often blurry, and organisms are not unitary. Positing an internal or external locus of control depends on having *some* extant division between a system and its environment, and often which system within the organism is posited varies with the phenomenon of interest. Candidates for an internal locus of control thus might be the genome, the brain, the circulatory system, the immune system, etc. External environments might include the organism's ecological niche, its conspecifics, its family, and so on. When we talk about internal versus external locus of control, we talk about whether factors internal or external to the relevant system drive the phenomenon of interest.

Let's return to our doll examples. The explanandum is the movement of the dolls. In the marionette case, it is the movements of the puppeteer and the strings that explains the way the doll moves. If the puppeteer moved even slightly differently, the marionette would move differently as well. In the doll, however, it is the movement of the motor, as well as the linkages to the doll's "limbs" that explains the doll's movement.

It is important to note that internalists are not committed to *never referencing* external factors in their explanations; nor are externalists committed to never referencing internal factors. What movements the marionette can perform will depend, to some degree, on the structure of the puppet's joints, what material it is made of, etc. The behavior of the doll will depend on whether anyone has turned it on. What is important, however, is that these references to factors not in the locus of control are not causally specific. For any token explanandum, the fine-grained control of that explanandum will be posited to reside in the locus of control – external for the marionette, internal for the doll. Or, to take an even simpler example: the radio will not work if I unplug it. But the presence or absence of electricity is a coarse-grained influence, compared to the manipulation of the dial.[1]

In a distributed locus of control, alternatively, internal and external factors are, at least in principle, equally specific in the grain of control they exert on the phenomenon, and thus manipulating either would change the outcome in a fine-grained way. Systems with distributed loci of control are harder to characterize, and uncontroversial examples are therefore a bit harder to come by. One explanandum that would be a candidate for a distributed-locus explanation is policy enactment in a representative democracy. Actors in the government make policy in such a system, but which actors are in office, as well as which policies they feel they can safely support, depend on public opinion. Another example may be commodities exchange in a regulated market. While the exchange of goods is driven by (amongst other things) supply and demand, these factors occasionally produce inefficiencies or exploitation, which are then regulated by governments. The exchange of commodities thus depends on both factors internal to and external to the market.

Vitally, in a distributed locus of control, the contributions of internal and external factors are independently characterizable. While both may exert fine-grained influence on the phenomenon, their contributions are not the *same*. The regulatory body does not exert the same control over commodities exchanges as firms do, for instance. Hence, any explanation of a system with a distributed locus of control must characterize independent, if interacting, causal forces internal and external to the system. The notion of a distributed locus of control is therefore different from the idea of a *holistic* system. Holists argue that there are no important divisions either within systems or between systems and their environments. This is because individual, internal components of the system make no specific causal contribution, aside from their interactions with other components and the environmental context (see Burnston, 2021 for a further exposition of holism). Since holists deny the division between a system and its environment, positing holism is logically distinct from positing any of internal, external, or distributed loci.

As a matter of fact, though, many holists in the debates we will discuss ally their views with externalism. This is because holists deny that internal components make any independent causal contribution to the phenomenon of interest. At best, on this kind of view, internal components are

---

[1] One way to put this difference is that an internal locus of control view will view external factors as *resources*, and vice versa for externalist views (Klein, 2018). Resources are non-specific, not-persistent causal influences on a phenomenon. The "mechanism" versus "resource" distinction, however, does not line up precisely with positing internal versus external loci of control, since there might be internal (e.g., metabolic) resources necessary even on an internalist conception.

coupled to or *driven by* external ones, and thus the fine-grained control in the system is due to external factors. We will thus consider holist views to be examples of externalism, and we will discuss several examples in the paper.

3. Internalism Versus Externalism in Evolutionary Biology

3.1 Internalism in the Modern Synthesis

The Modern Synthesis at the heart of 20th century evolutionary theory posited an internal locus of control for evolutionarily significant phenotypic change. On this view, the genotype of an individual produces its specific phenotype. Hence, differences in the genetics of different individuals drive phenotypic variations, which in turn are the basis of selection and evolutionary change. The Modern Synthesis thus combined a Darwinian perspective on evolutionary change with emerging knowledge of genetic mechanisms. Consider the following quote from Thomas Hunt Morgan:

> *All the genes are instrumental in producing each organ of the body. This may only mean that they all produce chemical substances essential for the normal course of development. If now one gene is changed so that it produces some substance different from that which it produced before, the end-result may be affected, and if the change affects one organ predominantly it may appear that one gene alone has produced this effect. In a strictly causal sense this is true, but the effect is produced only in conjunction with all the other genes.* (Morgan, 1926, page 302)

In this quote, Morgan evinces an internalist notion of the drivers of phenotypic variation. Note the implicit interventionist attitude taken towards genes – it is posited that changes to the genome are causally responsible for phenotypic outcomes. As such, explanations for variations in phenotype are to be given by citing causes in individuals' genes. The quote, importantly, does not evince a naïve atomism about genetic influence. It is admitted here that individual genes can influence specific traits at a variety of degrees of specificity, and further that an individual gene's contribution can only operate in conjunction with both other genes. Nonetheless, the locus of control for an individuals' traits is posited to be in its genes – the genetic makeup of the individual as a whole exerts specific causal influence on its phenotype.

The role of the genes is to determine the "normal course of development." This is a tacit admission that external factors can of course affect how phenotype comes out – removing oxygen from an aerobic organism's environment, for instance, will prevent it from developing. This is however a *non-specific* influence on evolved phenotypic variation, much akin to removing the power from the radio. In some cases, the phenotype reactivity to different environmental factors has been studied, but considered as a global and dynamic property of genotypes and proposed in the concept of *norm of reaction* (Sarkar, 1999). Also in the concept of norm of reaction however, fine-grained control over heritable changes resides in the individual's genes, because genes are causal factors that are differentially inherited across generations and control phenotypic expressions in different environments. Therefore, the locus of control for phenotypic variations is internal.

As molecular biology developed in the mid-20th century, the idea of an internal locus of control for traits was taken up in the famous "central dogma" of molecular biology. On this view, genetic 'information' is transmitted unidirectionally from genes (the sequence of nucleic acids in the DNA) to RNA to protein, and never in reverse. This shaped the molecular notion of genes as a conserved syntax -like sequence of chemical bases, the genome as a *program*, which encodes a set of instructions for development (Portin, 1993). On this view, an individual inherits a genetic program, encoded in the syntax of its genetic code – DNA has often been referred to as a *code of life* (Borek, 1969) or a *blueprint* (Gaffron, 1960). This view fits naturally with the internal locus of control, since such analogies posit fine-grained influence of genotype on phenotype. Fine-grained changes to the code/blueprint will result in fine-grained changes in the organism's traits.

Let's consider a classic example, which we will use to contrast the internalist view with externalism in the next section. During the industrial revolution in Great Britain, certain moths, *Biston betularia*, changed their colors. While the moths, prior to the industrial revolution, were light-colored and "peppered," air pollution from industrialization produced a layer of suit on the trees that peppered moths perched upon. Over time, a larger number of peppered moths began to exhibit a distinct phenotype, turning from peppered to almost solidly black.

On the internalist story, change occurs through random mutation in the genome. Peppered moths must have either already possessed different alleles for different colors (Cook & Saccheri, 2013), or undergone a mutation which produced, in some individuals, the black phenotype (Van't Hof et al., 2016). The new phenotype was then available for natural selection, which operated on it due to the camouflage advantage experienced by darker moths in ssoot-covered trees. So, the specific change in phenotype is caused by the specific change in genotype. Concomitant with this approach is a search for genetic modifications that are implicated in producing the phenotypic change. The environment applies selection pressure, but it does not influence the genetic variation that determines phenotypic outcomes. That is, the Modern Synthesis conception explains traits by positing an internal locus of control.

3.2. Critiques of the Internalist Conception

Major critiques of the Modern Synthesis, we suggest, are based on criticisms of the internal locus of control that it posits. Concomitantly, they are often pitched as critiques about the preformist notion of genes, arguing that phenotypes are in fact determined by the particular ways in which organisms interact with their environments. As such, critics of the Modern Synthesis propose as an alternative the idea that phenotypic variation has an external locus of control.

There are a variety of positions that fall under this rubric. For an early statement, consider the following quote:

Contrary to the neo-Darwinian view, we point out that the variations of the phenotype, on which natural selection could act, do not arise at random; they are produced by interactions between the organism and the environment during development (Ho & Saunders, 1979, page 1).

Building on an analogous proposal from Konrad Waddington (Waddington, 1977), the authors here state that phenotypic change is *not* due to random change in the genotype. Instead, it is driven by specific environmental variation. Change the individual's environment in particular ways, the quote suggests, and you will change its phenotype. Hence, the quote posits an external locus of control.

On the views proposed by externalists, the syntax-based concepts of genes and the genome as program are untenable. Critics have emphasized that the concept of "genetic program" does misrepresent the role of development and environmental contexts and promote a reductionistic view of evolution. On these views, genes are better described as participating to regulatory networks or as subroutines in the developmental operating systems (Falk, 1986). The genes concepts were thus progressively based on views including extended regulatory networks as to include epigenetic interactions and cellular products.

Externalist views thus focus on the causally specific influence on heritable phenotypic traits by environments. As noted above, externalists do not have to deny that genetic changes are causally relevant to variations in phenotype. In particular, externalists often posit that variations in genotype are *driven* by changes in environment, through epigenetic interactions (Jablonka & Lamb, 1989). What is important, however, is that the directionality of this process is distinct from the directionality on internalist views. Genes, on their own, do not produce specific phenotypic variation on externalist views. Instead, interactions with the environment both drives genetic expression and “unlock” extant genetic variation, causing it to produce specific phenotypic changes in specific environments (Hall, 1992). So, the locus of control for the variation is in external factors.

To see the contrast here, consider how an externalist would explain the peppered moth case. On this view, the interaction of peppered moths with their environments produces variation in their phenotype that was not already present. Again, this process may partially run through already extant genetic variation, but the genetic variation that was present is non-specific to black coloring. Black coloring is the result of specific interactions with an environment. This view, like its internalist alternatives, comes along with a methodological/explanatory prescription, namely to search for how environmental information is recognized by the organism and results in phenotypic plasticity (Eacock et al., 2019).

A variety of positions in evolutionary biology take up this alternative, externalist notion of the locus of control. On these views, it is an environment with a particular structure that explains heritable phenotypic variations through epigenetic regulations of genes, and not the organisms’ genes per se. The heritable changes in the phenotype produced by the environmental influences on the organisms (eco-phenotypes) can in turn be *stabilize*d across evolutionary time (Johnston & Gottlieb, 1990). Epigenetic interactions between organisms and environments can become standard and produce canalized developmental pathways leading to consistent phenotypes across individuals. For instance, the post-industrialization environment of the peppered moth could exert developmental/epigenetic effects on successive generations of moths, underlying the stability of the black coloring phenotype across generations.

Another position to discuss in this tradition, but one that requires some care in exposition, is developmental systems theory (DST) (Ford & Lerner, 1992; Oyama, 2000). DST proponents, like other critics of the Modern Synthesis, reject the idea of genes as privileged causal factor in development, as well as the idea of the genome as a program for producing phenotypes. DST theorists, though, often end up embracing a kind of *holism* about genetic/environment interactions (Deichmann, 2017; Sterelny et al., 1996) [but see (Griffiths & Stotz, 2018)]. On this view, there is one continuous systemic interaction between genetic and environmental features, and the unit of analysis is the developmental system. As we noted, holist positions are not the same as externalist ones, because of the emphasis on the organism, rather than on internal or external factors. However, there are aspects of DST that fit with an external locus of control. Specifically, in DST phenotypic change is often taken to be the result of continuous dynamic interaction between a system and its environment (Stotz, 2006). Hence environmental changes produce phenotypic change by coupling with internal factors in a non-decomposable way.

3.3. Contemporary Evolutionary Synthesis

We have argued that the genetic Darwinism of the Modern Synthesis was an internalist conception of the locus of control for evolved variations, and that criticisms of this view have taken an externalist position. We think that, in evolutionary biology, the field is trending towards a minimal agreement that the locus of control for evolved variation is distributed. On a distributed locus of control view, both genetic and environmental/developmental factors exhibit specific, non-redundant causal effects on heritable phenotypic changes during evolutionary processes. The phenotypic outcome for organisms is thus due to both kinds of factors.

On this view, genetic variation may be responsible for development of new traits, but the development of these traits requires specific contributions from developmental processes and environment. There are no *prime movens* or in principle directions of causal primacy. Genes are indeed heritable factors in producing phenotypes. But often, developmental and environmental factors will influence genes through epigenetic regulation, and traits will develop within, and take advantage of, environmental niches. The environment can thus influence heredity in a variety of ways. For example, the environmental influences (or different environmental niches) can affect genetic expression through epigenetic regulation, therefore varying the rate and type of heritable mutations.

In contemporary evolutionary biology, a variety of genes concepts have been proposed and are used in different experimental contexts and sub-disciplines. Genes are understood in a pluralist fashion (Brigandt, 2013) and both scientists and philosophers recognize that the concept of genes has evolved and will continue to evolve with scientific discoveries, explanatory practices and goals (Waters, 2006). However, no matter how genes are conceptualized, the recognized causal effects of genes reflect the distributed locus of control view we are defending here.

This point is well expressed by Waters in his “Pluralist interpretation of gene-centered biology” (Waters, 2006). While commenting on the criticisms of environmentalists and developmental system theorists to the gene-centrism and preformation of the Modern Synthesis, he states:

Biological research, much of it gene - centered, has indeed shown that inheritance is systematic and involves the interaction of genes, accessory molecules, cellular structures, and the surround. Furthermore, [critics have] a valid point about those who would want to identify nature with genes *and nurture with environment. What I want to emphasize is that this argument [ndr against causal privilege of genes] does not show, and apparently does not purport to show, that genes cannot be "properly contrasted" with environment in certain contexts.* (Ibidem, page 193)

What Waters is saying here is that, irrespective of the many definitions and conceptions of genes in various scientific fields and in philosophical reflection, scholars acknowledge that genes are only one of the many causal factors having effects on phenotypes and that it is important to disambiguate the specific causal role of the environment and of genes in producing inherited variations. As such, this claim is useful to capture the minimal agreement we have outlined here, which suggests that, rather than either an internalist or externalist view, the best approach posits a distributed locus.

As noted in section 2, it is vital on a distributed locus of control view that internal and external components be independently characterizable. Understanding, for instance, how epigenetic regulation works requires distinguishing the genetic component from the environmental factors that regulate it, and in turn understanding their interaction. While the result may be a kind of "reciprocal causation" (Laland et al., 2013), it is a reciprocal causation between functionally distinct elements (Buskell, 2019; Svensson, 2018).

Let's apply the distributed control perspective to the peppered moth example. This perspective suggests that there will be both internal and external factors that will contribute non-redundantly to the variation in phenotype. So, while it is important to understand if there are genetic variations specific to possession of the phenotype, this in no way undermines the need to look for environmental and epigenetic processes that also may be required. Both kinds of factors will affect the evolution of color distributions across the moth population. Indeed, many current models attempt to account for and differentiate the role of these distinct factors, as well as the reciprocal interactions between them, in a bidirectional eco-evolutionary feedback system (Alberti et al., 2017).

It is important to note the sense in which the kind of agreement we've outlined here is "minimal." What the minimal agreement suggests is that there is no *in principle* primary causal factor or causal directionality. This is compatible with different specific traits being more-or-less under the control of genetic or environmental factors. It is also compatible with a range of theoretical positions about the nature of evolutionary theorizing. Some, for instance proponents of the "Extended Evolutionary Synthesis," have suggested that the recognition of reciprocal causation is both an expansion and a foundational revision of Darwinian evolutionary theory (Laland et al., 2014). Others think that the recognition is compatible with and simply adds to traditional Darwinism – i.e., that the 'Extended' in 'Extended Synthesis' should be lowercase (Wray et al. 2014).

Finally, the notion of a distributed locus of control is also compatible with the methodological focus of a specific scientist or group being primarily on one kind of factor or another. Indeed, the distributed locus view explains the need for this differentiation, since both internal and external factors need to be independently characterizable on that view. So, the minimal agreement we've outlined does not refer to a monolithic evolutionary theory or methodology. What it does suggest is the recognition of the field that the domain is not structured in a way that supports either internalism or externalism about the locus of control. In the remainder of the paper, we suggest that a similar progress is needed for theoretical debates in cognitive science.

4. Locus of Control in Cognitive Science.

4.1. The Debate.

In the previous sections, we argued that traditional debates in evolutionary biology have been driven by disagreements about the locus of control for heritable phenotypic variation. We also suggested that a minimal agreement has emerged that the locus of control for evolved traits is, in fact, distributed. In this section we argue that debates in cognitive science are at a stage similar to the traditional debate in biology, and that they need to move in a similar direction. That is, theorists are debating between internal and external loci of control, when the consensus should be that the locus of control is distributed.

We further suggest that, in the cognitive science debates, the idea of mental and neural 'representations' should be viewed similarly to that of 'genes' in the biology debates. That is, it became a lightning rod for contention in the traditional debates, only to evolve to serve a more modest explanatory role in the agreement about distributed locus of control. Again, we suggest that a similar development is both possible and desirable in the cognitive science case.

In cognitive science, we presume that the phenomenon of interest is intelligent behavior (we offer no analysis of 'intelligent' here). From its foundation, "classical" cognitive science was dedicatedly internalist. Founded on the analogy to digital computers, classical cognitive science posited rich internal data structures underlying thought and behavior (Fodor, 1975; Newell & Simon, 1976). Mental processes were equated with manipulations of syntactic strings in a computational language according to internalized rules. The semantic interpretation of these strings is the foundation for the idea of a mental representation. Further, classical cognitive science viewed the relationship between the mind and the body in a similar way as the modern synthesis viewed the relationship between genes and traits – on the classical view, the mind encodes a set of instructions that the body follows, hence producing observable behavior.

On the classical view, factors internal to the organism's mind are the locus of control. Changing the mental 'program' being run by the organism – or, in its folk-psychological guise, changing the organism's beliefs and desires, which are presumed to be internal states – will exert fine-grained changes on the behavior of the subject. The idea of an internal program or "model" has also been foundational for cognitive neuroscience, which, traditionally, has taken as its goal the localization of the representations and mental functions posited in cognitive theories (Boone & Piccinini, 2016; Shea, 2018).

In the last 25 years, a variety of alternative research programs have developed, focusing on embodiment and dynamics as opposed to internal representational structures. We focus on the position often referred to as "enactivism" (Gallagher, 2017; Hutto et al., 2014; Thompson & Varela, 2001). This view is often pitched explicitly against the internalism of the classical position. Enactivists are fond of pointing out that, according to classical cognitive science, an organism could be a brain in a vat and their mind would be unchanged – since the locus control for their thoughts and behaviors is internal, the facts of their external environment are at best minimally relevant in determining what goes on in their mind. Enactivists propose that this view of the mind is radically false. The mind, they suggest, can only be understood as embedded in an environment; we misunderstand the mind by talking about it solely in terms of an internal program performed over mental representations. Instead, they posit that the very nature of the mind is determined by the environments with which it interacts.

In our reading, enactivism embraces an external locus of control for behavior, which is sometimes combined with a holism about mental systems. Concomitant to these moves is a view about the traditional explanatory posits and methodological aims of both classical cognitive science and cognitive neuroscience. In particular, enactivists systematically deny the existence and explanatory usefulness of mental representations, and they deny the possibility or desirability of functionally decomposing the brain. Rather, they suggest, one can only understand the mind and brain as holistic systems coupled to the environment. This view is externalist because it denies the independent contribution of factors internal to the mind, and thus claims that fine-grained control of behavior is due to the environment in which cognitive agents act.

As such, enactivism comes along with the standard set of methodological and theoretical prescriptions – it suggests limited utility to characterizing internal factors independently of the external ones with which they interact. Specifically, enactivists deny that cognitive science ever benefits from positing internal informational structures, even in cases such as memory or imagery. This methodological stricture applies to the brain as well. While enactivists admit that the brain makes *some* contribution to behavior, perhaps by synaptic or genetic changes in the case of memory (Hutto et al., 2014), or by physiological response to the environment during behavior (Gallagher, 2017), they insist that (i) there is no independent representational characterization of that contribution, and, concomitantly, (ii) the contribution can only be characterized as part of a dynamic coupling driven by the environment. In other words, the contribution of neural or genetic factors outside of environmental interaction is *non-specific*, in the sense outlined above. While some genetic or physiological processes may be necessary for memory or behavior, the explanation of their contributions depends asymmetrically on external factors.

We wish to note at the outset that we think enactivist approaches have made significant contributions to our understanding of the mind. We agree with enactivism that the internal locus of control posited by classical cognitive science is untenable. But we disagree with their externalist approach just as much. We suggest that, on a parallel with the emerging minimal agreement in evolutionary biology, cognitive and neuroscience need to embrace a distributed locus of control. And, while it is logically separable from this main point, we suggest that this makes room for the concept of mental and neural representation. Similar to how the concept of *gene* has evolved from a causally specific program to a contributing causal factor in a distributed locus of control, we suggest that the internal *aspect* of the distributed locus of control for cognition is plausibly characterized in representational terms.

There is no universally agreed upon set of conditions for what makes something a mental representation, although there are a number of proposals for what the "core," minimal notion of representation comprises (Bermudez, 2007; Piccinini, 2020, Ramsey, 2007; Rowlands, 2017; Rupert, 2018)). There have also been explicitly revisionary proposals how to construe mental representations outside of the classical picture (e.g., Goldman, 2012; Rowlands, 2009). Enactivists are fond of denying that these alternative proposals *really* offer an account of mental representations, rather than simply re-describing enactivist points in representational language (Gallagher, 2017). One might take this situation as sufficient reason to just abandon the notion of representation altogether – not as false *per se*, but as not theoretically useful (Rowlands, 2017).

In our view, there is no pre-theoretically agreeable set of necessary and sufficient conditions for what counts as a representation, and therefore the notion is likely to evolve with time. Moreover, there is nothing objectionable to this evolution so long as the particular theoretical commitments being undertaken by particular theorists are clear; indeed, this kind of evolution mirrors the evolution of the concept 'gene' and the interpretation of its causal role in evolutionary biology (Waters, 2006). We suggest that locating 'representations' amongst the internal locus of control is one productive way for the discussion to advance.

Some recent advancements from Rowlands (2017) are helpful in this regard. First, Rowlands suggests that at the core of the idea of representation are the intertwined ideas of *decouplability* and *normativity*. If a mental state or process is decoupled from the environment, then its function is understandable independently of its active interaction with aspects of that environment. Further, and partially *because* of the decouplability of the function, it has a *normative* aspect. That is, it can fulfill or not fulfill its function. Importantly, Rowlands notes that these are not offered as necessary and sufficient conditions – other theorists can and do add other notions, or attempt to abandon these. But they are particularly useful in our discussion, for two reasons: because they help clarify the opposition with enactivism, in arguing for decouplability where enactivists deny it, and because they presume no particular account of mental *content*. Particular notions of what mental content must amount to often inform debates about whether representation exists at all. But there are different accounts of what semantic content might be (Rupert, 2018), so it seems, in our view, unwise to insist on one as constitutive of representation.

Certain instances of learning, on our view, are clear evidence of the distributed locus of control (cf., Stotz & Allen, 2012). We argue that particular patterns in how agents learn new behaviors, shown at both the psychological and the neural level, evidence functioning of brain systems in a way that is not strictly definable as coupling with the environment. Enactivists are often vague about what they think learning does, saying that learning allows the agent to become attuned to new aspects of the environment, without representing it. We think this is insufficient – the outcome of learning is an internal structure that is not fully explicable in terms of coupling with specific environments. At the same time, however, they do not constitute programs or complete internal models of the environment. Both the nature of the learning process, and the success of the learned behavior, depend on a coordination between the learned internal/mental structure and the environmental or task context. Hence, the locus of control for behavior is distributed.

Here is how we think the story goes. In a learning context, agents develop internal mental structures and processes that track the structure of the task situation. Consider this as a kind of statistical learning about the task situation, which is not necessarily nor strictly determined top-down by agents' prior knowledge or assumptions (Burnston, 2021a, 2021c). Hence, the nature of the internal process depends, etiologically, on the structure of the environment. A different environment would produce a different learning outcome. However, once the structure is learned, the structure enables generalization to novel examples and instances. Change the internal structure, and you would get a different pattern of generalization, and a different range of environments in which the behavior would be successful. So, both internal and external factors exhibit fine-grained causal contributions to successful behavior, but those are not the same contributions.

We suggest that the internal causal factors here, which we will refer to as "internal structures" to avoid begging the question, are good candidates for a modern conception of mental representation, because they meet decouplability and normativity. Decouplability is evidenced by the fact that the contribution of internal structures is not fully explained by interacting with specific environments. It is only by talking about the internalized structure that one can explain patterns of generalization. But normativity is explained by the fact that whatever environment is generalized to must share a structure with the learned environment in order for behavior to be successful. In what remains of the paper, we discuss examples meeting this description of internal structure at both the psychological and neural levels.

5.2. Psychological Examples

We start with a very simple example about how environmental factors contribute to the allocation of attention. By "factors" here we simply mean aspects of the environment; this can include, for instance, regularities amongst the kind of stimuli the organism encounters, or in the kinds of rewards available in the environment. A simple kind of investigation in this vein looks for how regular structures in the environment are used to shape attentional patterns. For instance, in a study by (Zang et al., 2016), subjects were instructed to look for a shape amongst distractors. In some experimental groups, the shapes were distributed around a task-irrelevant display structure, such as a large cube. Despite the task-irrelevance of this structure, subjects learn to use it to guide their attentional search. When it is removed in a subsequent task epoch, subjects' search success goes down, even if the layout of shapes and distractors remains the same.

We suggest that his simple case exhibits the structure of a distributed locus of control. What subjects learn depends on the task environment – if you changed the shape of the cube, for instance, subjects would learn a different search pattern. Once learned, however, the internalized structure shapes behavior, even in different environments (e.g., the ones without the cube). If you changed the internal structure, the search behavior would be different. The success of the behavior requires that the two independently characterizable components – the environment and the learned structure – to be aligned. So, while the internal component is decouplable from the environment, it needs to be employed in the right circumstances for behavioral success, hence underlying normativity.

Even clearer examples come in cases where generalization is explicitly investigated. In studies of motor control, for instance, subjects learn a task situation and then perform actions in a range of conditions to test what they've learned (Wolpert & Flanagan, 2016). These studies employ a motor control element and a visual feedback element. So, subjects might have to move a joystick, which controls a visually presented cursor, until the cursor reaches a target. In some conditions, the visual feedback is "perturbed" – the cursor, for instance, may move to the left or right – and subjects have to adjust to the perturbation by manipulating the joystick. Learning is shown by quicker and more efficient responses to perturbations.

A huge range of studies have shown that subjects learn to respond not only to individual perturbations, but to ranges of perturbations, which subsequently allows them to generalize (for a summary, see Burnston, 2021). So, if the perturbations occur across a range of angles and magnitudes, subjects will not only learn to respond to any of those perturbations, but will respond to a novel perturbation just as well, if it falls in the range.

Here we see the same kind of explanatory structure playing out. Subjects' learned abilities depend on a structure in the environment – in this case, the variational structure in the perturbations they experience. However, they also internalize that structure in a way that allows for generalization. It is thus not coupling to specific environments that determines behavior, but an internalization of task structure. Again, while a different task structure would produce a different internal structure, this does not mean that the internal structure is not an independent causal contributor. Of course, successful generalized behavior depends upon the new task condition falling within the generalized range. So, the locus of control is distributed, and the internal component is both decouplable and underlies normativity.

A last psychological example comes from studies of categorical perception (Goldstone et al., 2017; for in depth philosophical analysis of these an similar cases see Burnston, 2022, forthcoming). Studies in this field start with construction of a "space" of examples, such as the one shown below. The space is constructed by "morphing" two pairs of faces towards each other, generating new examples along the way. Each face is thus a combination of the four 'parents'. In the studies, experimenters determine what the categories will be. In the examples below, 'A faces' are on the left side of the (arbitrary) line in the middle, and 'B faces' are on the right. Subjects are trained with feedback, and after training become good at immediately recognizing a face as an A or a B. This involves discriminating along the horizontal "X axis" in the space.

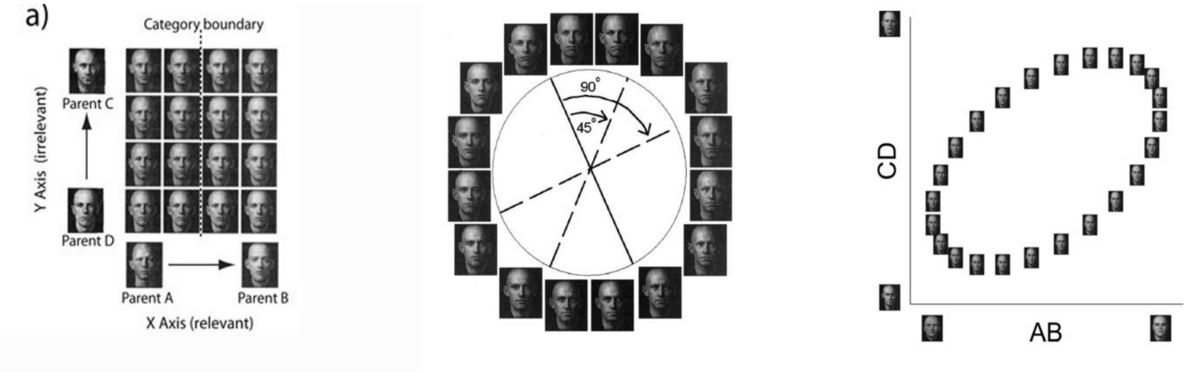

Figure 1, exemplar spaces. Left panel from Folstein et al. (2012). Middle panel from Goldstone & Steyvers (2001). Right panel from Jones and Goldstone (2013).

Importantly, subjects then generalize this learned structure in a variety of ways. For instance, if a new space of faces is constructed by keeping the parents comprising the X axis the same, while pairing them with a new set of parents, Z, subjects can still recognize the faces (Goldstone & Steyvers, 2001). That is, they learn to recognize the "X" pattern of variation even in new examples. Other kinds of generalization are also relevant. For instance, if in a subsequent task epoch the category boundary "turns" 90 degrees, subjects can transfer their ability almost immediately. But such a transfer is not shown if the category boundary turns 45 degrees (see middle panel, Figure 1). This is explained by the posit that subjects not only learn the dimension that defines the category, but the dimension orthogonal to it, as a way of fully parsing the space.

Importantly, there is no particular set of exemplars that is necessary to produce this result. Subjects can be trained on a range of exemplar sets, so long as those exemplars vary in the way that is relevant for subsequent category recognition (Jones & Goldstone, 2013). This is shown in the middle and right panels of Figure 1. Again, we have the same explanatory structure we have discussed above. What structure subjects learn depends on the variation that is present in the exemplar set and the feedback they receive. Once they have learned that structure, they can generalize it to new instances. But successful generalization requires the new instances to have the right kind of structure.

5.3. Neural Dynamics.

Enactivists regularly espouse a dynamic approach to the brain. Rather than thinking of the brain in terms of a functionally decomposable, information-manipulating system, they argue, we should understand it as dynamically coupled to the environment (Silberstein & Chemero, 2013). This claim rarely is pursued at anything other than an abstract level, however, and historical debates about the relationship between representation and dynamics have similarly taken place at a far remove from neuroscientific practice (see, e.g., Bechtel, 1998; for a more up-to-date discussion see Shagrir, 2012). There is no doubt that neural dynamics are vitally important for understanding brain function, but we should ask whether that fact is best read as supporting enactivism. In this section, we suggest that investigations into neural dynamics are in fact supportive of a distributed locus of control.

We will focus specifically on frontal cortex, a part of the brain that is involved in organizing complex behavior. Importantly, recent physiological and modeling studies of this area suggest that function is widely distributed, and that individual units in the system are multifunctional. Particular variables cause physiological responses in widely distributed groups of cells, and individual cells are "multiplexed," showing significant physiological responses to an array of task-related variables (Rigotti et al., 2013). Moreover, the same cells respond to distinct variables in distinct conditions, suggesting that function is highly context-sensitive. The functionality of a system like this is often determinable only at the population level, and that functionality consists in patterns of dynamic activity over time. Despite this, we think, studies in this area do not support an externalist view.

Instead, a vital part of understanding these systems involved *decomposing* the dynamics into distinct regimes that correspond to variations in task context. Again, learning is an important part of the story. We discuss only one highly-cited example here, but we take this as an exemplar for how to understand neural dynamics on a distributed locus of control-based view.

Mante et al. (2013) attempted to discern the function of a distributed population of prefrontal cortical cells in complex task situations. Their question was the following: how can a prefrontal population sensitive to multiple task-relevant factors change *which* factor should drive behavior in a particular context? To analyze this, they modified a traditional "dot-motion" task paradigm. In traditional dot motion studies, a subject, in this case a monkey, views an array of moving dots and has to judge the direction of their predominant motion. With no correlation in the motion of individual dots, there is no predominant motion. With small degrees of correlation monkeys remain close to chance at detecting it, but at greater degrees of correlation they can generally report the direction in which the motion of the dots is correlated.

Mante et al.'s modification was to also vary *color* along with motion. So, the color of the dots could be either red or green. The level of predominance – i.e., the proportion of the colors – could vary just as the degree of correlated motion did, and across trials both the color and motion predominance varied independently. When given a context cue telling them which was the relevant stimulus parameter, monkeys would indicate their choice via a saccade to the right or to the left. The results showed that in the color context, the monkeys could accurately judge the predominance of color independently of motion, and vice versa in the motion context. The question was how their brains separated and organized the information.

Importantly, widely dispersed prefrontal cell populations showed overlapping physiological selectivity for color and motion – they were distributed and multiplexed in the ways described above. Equally importantly, color information *continued to influence* physiological responses in this population in the motion-judging context, and vice-versa. So, the story is not a simple one of the monkey attending to the output of a distinct, independent color response in the color context, and a separate, independent motion response in the motion context. Somehow, the very same population that was responding to the task parameters extracted the relevant information in the correct context.

The full story of Mante et al.'s explanation for how this works is complex, but here are the basics (See Burnston, 2021c for a fuller description). First, they analyzed population responses in terms of *task axes,* represented in a principal components space. Principal components analysis (PCA) measures covariation in a high-dimensional system, in this case the physiological responses of a large population of individual neurons, in terms of orthogonal dimensions, small numbers of which are sufficient to describe the variation of the entire population. Mante et al. performed PCA, then constructed a description of the populations' specific responses to variation in motion, color, and context in terms of those components. That is, the population's response to each task parameter was construed as a particular subset of the overall population response. Lastly, they analyzed the axis of *choice* – i.e., which behavioral indication the monkey made as indicated by the direction of their saccade response. The outcome of the initial analysis was that the task context changed the relationship between the task parameters and choice. In effect, in motion contexts motion information would vary the population along the choice axis, while color information wouldn't, and the opposite in the color context.

To analyze this further, they trained a recurrent neural network to mimic the population response, and analyzed the dynamics of this network. They showed that the system exhibited two distinct "line attractors," corresponding to the motion context and the color context, and that the context cue would push the system into one or the other. On the line attractor for the motion context, motion information would drive the choice, and color would drive the choice on the attractor for the color context. Details aside, what is important here is that the system learned to *modify its own dynamics* given the task context. What the system does is learn to implement different dynamic regimes for the same set of stimuli, as the task context varies. Learning in this case allows the system to extract the right information from the environment in the right context.

We suggest that the study shows decouplability *in the dynamics* of the neural responses, and that this case exhibits the same pattern of explanation as the examples in the previous subsection. In particular, the nature of the learning depends on the environment in which learning occurs – it is the fact that color and motion vary in the environment that leads the PFC to implement selectivity for those variables. If you varied the training history, the system would track something different. What learning does, however, is establish a way of using or manipulating those informational responses that is functional for the task context. The task requires separating motion from color information, and employing the right information in the right context. The solution that the system hits upon, if the analysis is correct, is the development of distinct line attractors, such that the context cue pushes the system to the appropriate one for the task context.

At the beginning of the training, monkeys are not able to extract the relevant information. After the training, they are. What explains the difference is the way that information is organized internally so as to drive population dynamics. So, we have the same schema proposed in the previous section. Environmental variation and training history determine what is learned, but once the learning has taken place, the learned structure can generalize – the monkey's frontal cortex can respond, presumably, in a similar way to a degree of correlated motion it hasn't seen before, for instance. Of course, as we saw in the previous section, this generalization will only work so long as the task situation is appropriate. The ability to separate motion from color will not necessarily carry over to the ability to separate, say, timbre from pitch. While decoupling occurs, the success of behaviors guided by the internal changes will depend on novel task environments having a similar enough structure to those learned. Given the specific causal contributions of both internal and external components, even neural dynamics are best understood in terms of a distributed locus of control.

5. Conclusion.

We have argued that theoretical debates in both evolutionary biology and cognitive science have, at least implicitly, taken place between theorists positing different loci of control. Internalists place the locus of control internal to the (relevant subsystem of) the organism. Externalists place the locus of control outside of the (relevant subsystem of) the organism. We have contended that a productive trend is emerging in evolutionary biology which posits instead a distributed locus of control. Our conclusion in this paper is a normative one for debates in theoretical cognitive science: we should move beyond the internalism/externalism dichotomy by positing a distributed locus of control, and we should reconstrue key notions such as 'mental representation' along those lines.

In closing, we stress two further advantages of positing a distributed locus of control, both of which stem from the fact that it is an *explanatory strategy*. It suggests that a phenomenon of interest will be best explained as the result of specific, non-redundant contributions from both internal and external components. Read as an explanatory strategy, positing a distributed locus of retains a large amount of both methodological and metaphysical *flexibility*.

A view based on a distributed locus of control is *methodologically* flexible because it suggests that both internal and external components need to be independently identified and characterized. As such, if particular researchers (say, as in our case of the peppered moth in section 3) are primarily interested in a particular internal or external component, a distributed locus of control view supports their primary methodological focus on that component. What it denies, however, is that that methodological focus comes along with *explanatory* primacy. The necessity of individuating and studying both internal and external components, on our view, does not equate to situating the locus of control with one or the other. In cognitive science, a distributed locus of control view thus supports the investigation of information processing systems within the mind/brain, as well as views of cognition that focus on the cultural or material contributions to cognition (see, e.g., Heersmink & Sutton, 2020; Menary, 2015; Stotz, 2010).

A view based on a distributed locus of control is *metaphysically* flexible because it is ultimately neutral with how to draw boundaries around the phenomenon of interest. What it requires is functional decomposability – i.e., that *components* have distinct causal contributions. This is compatible with a metaphysical view on which biological mechanisms extend into the environment, or even one on which it is ultimately processes, and not mechanisms, that are at work (so long as the processes themselves are decomposable (Nicholson & Dupré, 2018; Roe & Baumgaertner, 2017)(Author's paper forthcoming). It is also compatible with a view of the mind on which the mind itself extends into the body or the environment (Kirchhoff, 2014; Sprevak, 2019). So long as the explanatory suggestions are respected, a distributed locus of control view can accommodate each of these alternatives.

Hence, we suggest, views based on distributed loci produce the best overall result for theorizing in cognitive science. They avoid falsely locating the locus of control either internally or externally to the system. They support a range of methodological and theoretical approaches. And they allow for those approaches to be pursued independently of broader metaphysical issues. This is the best approach for a functioning science of organism and mind.

References

Alberti, M., Marzluff, J., & Hunt, V. M. (2017). Urban driven phenotypic changes: Empirical observations and theoretical implications for eco-evolutionary feedback. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1712), 20160029. https://doi.org/10.1098/rstb.2016.0029

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, *22*(3), 295–317. https://doi.org/10.1016/S0364-0213(99)80042-1

Bermúdez, J. L. (2007). *Thinking without words*: Oxford University Press.

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, *193*(5), 1509–1534.

Borek, E. (1969). The code of life. *The Code of Life.* https://www.cabdirect.org/cabdirect/abstract/19700104440

Brigandt, I. (2013). Explanation in Biology: Reduction, Pluralism, and Explanatory Aims. *Science & Education*, *22*(1), 69–91. https://doi.org/10.1007/s11191-011-9350-7

Burnston, D. C. (2021a). Anti-intellectualism for the learning and employment of skill. *Review of Philosophy and Psychology*, *12*(3), 507–526. https://doi.org/10.1007/s13164-020-00506-5

Burnston, D. C. (2021b). Getting Over Atomism: Functional Decomposition in Complex Neural Systems. *British Journal for the Philosophy of Science*, *72*(3), 743–772. https://doi.org/10.1093/bjps/axz039

Burnston, D. C. (2021c). Contents, vehicles, and complex data analysis in neuroscience. *Synthese*, *199*(1), 1617–1639. https://doi.org/10.1007/s11229-020-02831-9

Burnston, D. C. (2021d). Bayes, predictive processing, and the cognitive architecture of motor control. *Consciousness and cognition, 96*. doi:10.1016/j.concog.2021.103218

Burnston, D. C. (2022.) How to think about higher-level perceptual contents. *Mind and Language*. doi: https://doi.org/10.1111/mila.12446.

Burnston, D. C. (forthcoming). Perceptual learning, categorical perception, and cognitive permeation. *Dialectica*.

Buskell, A. (2019). Reciprocal Causation and the Extended Evolutionary Synthesis. *Biological Theory*, *14*(4), 267–279. https://doi.org/10.1007/s13752-019-00325-7

Cook, L. M., & Saccheri, I. J. (2013). The peppered moth and industrial melanism: Evolution of a natural selection case study. *Heredity*, *110*(3), 207–212. https://doi.org/10.1038/hdy.2012.92

Deichmann, U. (2017). Hierarchy, determinism, and specificity in theories of development and evolution. *History and Philosophy of the Life Sciences*, *39*(4), 33. https://doi.org/10.1007/s40656-017-0160-3

Eacock, A., Rowland, H. M., van't Hof, A. E., Yung, C. J., Edmonds, N., & Saccheri, I. J. (2019). Adaptive colour change and background choice behaviour in peppered moth caterpillars is mediated by extraocular photoreception. *Communications Biology*, *2*(1), 1–8. https://doi.org/10.1038/s42003-019-0502-7

Falk, R. (1986). What is a gene? *Studies in History and Philosophy of Science Part A*, *17*(2), 133–173. https://doi.org/10.1016/0039-3681(86)90024-5

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

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object representations: Not all morphspaces stretch alike. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 807–820. https://doi.org/10.1037/a0025836

Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory: An integrative approach* (pp. xi, 259). Sage Publications, Inc.

Gaffron, H. (1960). The origin of life. *Perspectives in Biology and Medicine*, *3*, 163–212. https://doi.org/10.1353/pbm.1960.0001

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind*. Oxford University Press.

Goldman, A. I. (2012). A Moderate Approach to Embodied Cognitive Science. *Review of Philosophy and Psychology*, *3*(1), 71–88. https://doi.org/10.1007/s13164-012-0089-0

Goldstone, R. L., Rogosky, B. J., Pevtzow, R., & Blair, M. (2017). The construction of perceptual and semantic features during category learning. In *Handbook of categorization in cognitive science, 2nd ed* (pp. 851–882). Elsevier Academic Press. https://doi.org/10.1016/B978-0-08-101107-2.00034-8

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139. https://doi.org/10.1037/0096-3445.130.1.116

Griffiths, P. E., & Stotz, K. (2018). Developmental Systems Theory as a Process Theory. In D. J. Nicholson & J. Dupre (Eds.), *Everything Flows: Towards a Processual Philosophy of Biology* (pp. 225–245). Oxford & New York: Oxford University Press.

Hall, B. K. (1992). Genetics, epigenetics and environment. In B. K. Hall (Ed.), *Evolutionary Developmental Biology* (pp. 150–182). Springer Netherlands. https://doi.org/10.1007/978-94-015-7926-1_9

Heersmink, R., & Sutton, J. (2020). Cognition and the Web: Extended, Transactive, or Scaffolded? *Erkenntnis*, *85*(1), 139–164. https://doi.org/10.1007/s10670-018-0022-8

Ho, M. W., & Saunders, P. T. (1979). Beyond neo-Darwinism—An epigenetic approach to evolution. *Journal of Theoretical Biology*, *78*(4), 573–591. https://doi.org/10.1016/0022-5193(79)90191-7

Hutto, D. D., Kirchhoff, M. D., & Myin, E. (2014). Extensive enactivism: Why keep it all in? *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00706

Jablonka, E., & Lamb, M. J. (1989). The inheritance of acquired epigenetic variations. *Journal of Theoretical Biology*, *139*(1), 69–83. https://doi.org/10.1016/S0022-5193(89)80058-X

Johnston, T. D., & Gottlieb, G. (1990). Neophenogenesis: A developmental theory of phenotypic evolution. *Journal of Theoretical Biology*, *147*(4), 471–495. https://doi.org/10.1016/S0022-5193(05)80260-7

Jones, M., & Goldstone, R. L. (2013). The structure of integral dimensions: Contrasting topological and Cartesian representations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 111–132. https://doi.org/10.1037/a0029059

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006). *Scientific Pluralism*. U of Minnesota Press.

Kirchhoff, M. (2014). Extended cognition & constitution: Re-evaluating the constitutive claim of extended cognition. *Philosophical Psychology*, *27*(2), 258–283. https://doi.org/10.1080/09515089.2012.724394

Klein, C. (2017). Brain regions as difference-makers. *Philosophical Psychology*, *30*(1–2), 1–20. https://doi.org/10.1080/09515089.2016.1253053

Klein, C. (2018). Mechanisms, resources, and background conditions. *Biology & Philosophy*, *33*(5), 36. https://doi.org/10.1007/s10539-018-9646-y

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*(5), 719–745. https://doi.org/10.1007/s10539-012-9335-1

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E., Futuyma, D. J., Lenski, R. E., Mackay, T. F. C., Schluter, D., & Strassmann, J. E. (2014). Does evolutionary theory need a rethink? *Nature*, *514*(7521), 161–164. https://doi.org/10.1038/514161a

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–97.

Menary, R. (2015). Mathematical Cognition: A Case of Enculturation. In T. M. J. Windt (Ed.), *Open Mind*. https://philarchive.org/rec/MENMCA-2

Morgan, T. H. (1926). Genetics and the Physiology of Development. *The American Naturalist*, *60*(671), 489–515.

Newell, A., & Simon, H. A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Acm*, *19*, 113–126.

Nicholson, D. J., & Dupré, J. (2018). *Everything Flows: Towards a Processual Philosophy of Biology*. Oxford University Press. https://philarchive.org/rec/NICEFT-3

Noble, D. (2008). Genes and causation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1878), 3001–3015. https://doi.org/10.1098/rsta.2008.0086

Oyama, S. (2000). *THE ONTOGENY OF INFORMATION: DEVELOPMENTAL SYSTEMS AND EVOLUTION*. 213260. https://repository.library.georgetown.edu/handle/10822/546678

Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*: Oxford University Press.

Portin, P. (1993). The Concept of the Gene: Short History and Present Status. *The Quarterly Review of Biology*, *68*(2), 173–223. https://doi.org/10.1086/418039

Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge University Press.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590.

Roe, S. M., & Baumgaertner, B. (2017). Extended Mechanistic Explanations: Expanding the Current Mechanistic Conception to Include More Complex Biological Systems. *Journal for General Philosophy of Science*, *48*(4), 517–534. https://doi.org/10.1007/s10838-016-9356-6

Rowlands, M. (2009). Situated Representation. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 117–133). Cambridge: Cambridge University Press.

Rowlands, M. (2017). Arguing about representation. *Synthese*, *194*(11), 4215–4232. https://doi.org/10.1007/s11229-014-0646-4

Rupert, R. D. (2018). Representation and mental representation. *Philosophical Explorations*, *21*(2), 204–225. https://doi.org/10.1080/13869795.2018.1477979

Sarkar, S. (1999). From the Reaktionsnorm to the Adaptive Norm: The Norm of Reaction, 1909–1960. *Biology and Philosophy*, *14*(2), 235–252. https://doi.org/10.1023/A:1006690502648

Shagrir, O. (2012). Structural Representations and the Brain. *British Journal for the Philosophy of Science*, *63*(3), 519–545. https://doi.org/10.1093/bjps/axr038

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press. https://philarchive.org/rec/SHERIC

Silberstein, M., & Chemero, A. (2013). Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences. *Philosophy of Science*, *80*(5), 958–970. https://doi.org/10.1086/674533

Sprevak, M. (2019). Extended cognition. *Routledge Encyclopedia of Philosophy*. https://doi.org/10.4324/9780415249126-V049-1

Sterelny, K., Smith, K. C., & Dickison, M. (1996). The Extended Replicator. *Biology and Philosophy*, *11*(3), 377–403. https://doi.org/10.1007/BF00128788

Stotz, K. (2006). With ‘Genes’ Like That, Who Needs an Environment? Postgenomics’s Argument for the ‘Ontogeny of Information.’ *Philosophy of Science*, *73*(5), 905–917. https://doi.org/10.1086/518748

Stotz, K. (2010). Human Nature and Cognitive?Developmental Niche Construction. *Phenomenology and the Cognitive Sciences*, *9*(4), 483–501. https://doi.org/10.1007/s11097-010-9178-7

Stotz, K., & Allen, C. (2012). From Cell-Surface Receptors to Higher Learning: A Whole World of Experience. In K. S. Plaisance & T. A. C. Reydon (Eds.), *Philosophy of Behavioral Biology* (pp. 85–123). Springer Netherlands. https://doi.org/10.1007/978-94-007-1951-4_5

Svensson, E. I. (2018). On Reciprocal Causation in the Evolutionary Process. *Evolutionary Biology*, *45*(1), 1–14. https://doi.org/10.1007/s11692-017-9431-x

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, *5*(10), 418–425. https://doi.org/10.1016/S1364-6613(00)01750-2

Van't Hof, A. E., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C., & Saccheri, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, *534*(7605), 102–105. https://doi.org/10.1038/nature17951

Waddington, C. H. (1977). The Evolution of an Evolutionist. *Journal of the History of Biology*, *10*(2), 369–370.

Wolpert, D. M., & Flanagan, J. R. (2016). Computations underlying sensorimotor learning. *Current Opinion in Neurobiology*, *37*, 7–11. https://doi.org/10.1016/j.conb.2015.12.003

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287–318. https://doi.org/10.1007/s10539-010-9200-z

Zang, X., Geyer, T., Assumpção, L., Müller, H. J., & Shi, Z. (2016). From Foreground to Background: How Task-Neutral Context Influences Contextual Cueing of Visual Search. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00852