Is a Nervous System Necessary for Learning?

José E. Burgos¹

Published online: 5 November 2018

© Association for Behavior Analysis International 2018

Abstract

In this article, I propose some elements for a conceptual foundation for a negative answer to the titular question, based on a historical conceptual analysis of some definitions of “learning” in the specialized literature. I intend such a foundation to include learning in living organisms as well as inorganic machines. After analyzing several behavioral and nonbehavioral definitions, I argue that although most of the former favor a negative answer, they tend to be restricted to living organisms and thus exclude inorganic machine learning. They also face the yet-unresolved issue of behavioral silence, which makes behavior not defining of learning. Some nonbehavioral neurobiological definitions favor an affirmative, others a negative answer, but still exclude inorganic machines. Nonneurobiological definitions are more suitable, but they commit us to some form of computationalism (Turing machine or connectionist) about learning, which is premature. I thus propose elements for an alternative definition of “learning” without such commitment. The elements are elaborations of the notions of learning as a kind of causal interaction between causal stochastic environmental and internal processes, and minimal learner as a kind of abstract system that shares certain internal structural and functional features with animals, spinal vertebrates, bacteria, plants, and inorganic machines.

Keywords Definitions · Concepts · “Learning” · “Behavior” · Machines · Internal mechanisms · Causal processes · Learners

Some readers might dismiss the titular question (Q) as obviously answerable in the affirmative and, hence, superfluous. In this article, I argue that a negative answer is not as mistaken as some might think. In fact, it is a better answer, because it provides an opportunity to search for a broader concept of learning that fosters links to other fields and, in this way, stimulates novel collaborative research. My aim here is to begin this

✉ José E. Burgos
jburgos@cucba.udg.mx

¹ Centro de Estudios e Investigaciones en Comportamiento, Universidad de Guadalajara, 180 Fco. De Quevedo, Arcos Vallarta, 44130 Guadalajara, Mexico

search. I will leave it unfinished, as it is a long, involved, uncertain, daunting task, as will soon become evident. I shall only seek a few elements that could contribute to an alternative definition of “learning” as a conceptual foundation for a negative answer to Q. My ultimate goal is conceptual clarification as integral to scientific progress (see Laudan, 1977).

The search is heavily conceptual, because much hinges on the concept of learning, or definition of “learning” (I shall treat the two equivalently). As Bunge and Ardila (1987) put it: “. . . the attribution of a learning ability . . . depends critically on the definition of ‘learning’” (p. 28). There are multiple definitions of “learning.” Some favor an affirmative answer to Q, others a negative answer. A negative answer is one that makes these existential statements true: “Some learners do not have a nervous system” equivalent to “Not all learners have a nervous system,” where by “learner” I mean a system with certain structural and functional features (discussed in the last section). An affirmative answer to Q is one that makes these universal statements true: “All learners have a nervous system” equivalent to “Only things with nervous systems are learners.”

Evidently, the existential and universal statements are mutually exclusive. In either case, I take the truth or falsity of a given statement as literal. In particular, if the existential statements, which favor a negative answer to Q, are true, they are *literally* true. In this sense, the present analysis is no mere semantic analysis of the meaning of “learning,” but has a strong ontological import, with implications for the nature of learning. I thus take definitions of “learning” to be expressions of concepts of learning as views about the nature of learning.

Such ontological import is apparent in several discussions about whether certain kinds of individuals can learn. Invertebrates have been widely discussed in this regard. Here is an illuminating example (I give more at the beginning of the first section):

Can invertebrates learn? Few other questions in comparative psychology have engendered such heated and passionate discussions as has this one. Quite understandably, the question cannot be answered adequately until some definition of the term “learning” is universally agreed upon, a most unlikely event at best. (McConnell, 1966, p. 109)

This quotation is echoed in Bunge and Ardila’s (1987) claim that an answer to Q hinges on the definition of “learning.” The presence of multiple definitions of “learning” also supports the assessment that a universally agreed-upon definition is unlikely: Here we are, over half a century later without an agreed-upon definition of “learning,” with most definitions largely unchanged. I also take the “heated and passionate” character attributed to these discussions in the above quotation as indicative of high stakes, and few stakes in science are higher than truth or falsity. If we follow most scientists and philosophers in adopting a correspondence theory of truth (I do), such discussions are about the nature of what exists. Here I will assume that whether a nervous system is necessary for learning is tightly intertwined with the nature of learning (as expressed by definitions of “learning”).

Still, invertebrates constitute too broad a category. For a long time now, there has been consensus on attributing a learning ability to insects and spiders (e.g., see Carew & Sahley, 1986; Dukas, 2009, 2018; Jakob, Skow, & Long, 2011; Matthews &

Matthews, 2010; Papaj & Lewis, 1993), which are invertebrates. Insects, in particular, perhaps the largest invertebrate animal class, are also widely regarded to have a central nervous system in the standard technical neuroscientific sense.

The strictest sense refers to a whole nervous system, central and peripheral, but authors are rarely if ever this strict, especially regarding the necessity of a nervous system for learning. They thus often use the expression “nervous system” more vaguely, to refer to a central nervous system (a brain connected to a spinal cord in vertebrates, or to a ventral nerve cord in some invertebrates, especially insects), a brain, or some major anatomical part of it. I will follow this practice here.

A brain, as traditionally conceived in neurobiology, is an organ consisting of glia (nonexcitable satellite cells that protect and nourish neurons) and various types of neurons interconnected through synapses. Of course, a nervous system is not a structurally static organ, but constantly and naturally loses neurons and glia, and gains glia (too many, in cases of glioblastoma). The line between a complete and an incomplete nervous system (and organism) is fuzzy.

Learning in relatively *whole* invertebrates and vertebrates entails an affirmative answer to Q (under the above logic), insofar as whole animals are made of whole nervous (and digestive, skeletal, circulatory, muscular, etc.) systems. However, many if not most living organisms that lack nervous systems, most naturally, others artificially, have also being attributed a learning ability. Here is a partial but representative list:

- Spinal (decerebrated) vertebrates (e.g., Chopin & Buerger, 1976; Grau & Joynes, 2005; Shurrager & Culler, 1940)
- Paramecia (e.g., Armus, Montgomery, & Jellison, 2006; Hennessey, Rucker, & McDiarmid, 1979; cf. Wichterman, 1986, pp. 237–238)
- Bacteria (e.g., Alon, Surette, Barkai, & Leibler, 1999; Yi, Huang, Simon, & Doyle, 2000)
- Plants (e.g., Applewhite, 1972; Gagliano, Vyazovskiy, Borbély, Grimonprez, & Depczynski, 2016)

All of this is *organic* in that its structure and function depend intrinsically on molecules of one or more carbon atoms covalently linked to atoms of other elements, mostly hydrogen, oxygen, and/or nitrogen, especially protein synthesis, genetic as well as epigenetic. Unless otherwise indicated, by “nervous system,” I mean “organic central nervous system” in this sense of the term “organic.”

The possibility of learning in spinal vertebrates favors a negative answer to Q, because they lack a major part of a central nervous system (a cortico-spinal link). The rest of the cases favor the same answer but more drastically, by the complete natural absence of a nervous system. If indeed these things can learn, this would suffice for a negative answer to Q. But I have a more ambitious goal: to include *inorganic machines* (e.g., Alpaydin, 2016; Gori, 2018; Langley, 1996; Mitchell, 1997; Simon, 1983). By inorganic machine learning (IML) I mean that inorganic machines learn literally, not figuratively.

IML thus viewed represents the strongest case for a negative answer to Q and makes this statement true: Some inorganic things can learn. This statement favors a negative answer to a more general question: Is organic life necessary for learning? The possibility of a negative answer is intriguing and I will take it seriously, because it could

expand our understanding of learning beyond organic life and, in this way, allow for a better understanding of organic learning. I thus will seek a conceptual foundation for IML (and hence a negative answer to Q) that also includes organic learning. My search is justified in its own right as a hypothesis about the nature of learning, not as a means to benefit the study of animal learning, although I do not rule out the possibility of some benefit either.

Figure 1 depicts the aforementioned families of potential learners, as a *fuzzy* partition (dashed curves) of the set of learners, also depicted as a fuzzy set, to allow for conceptual flexibility. The proper relation of a particular individual to one of these families, then, is fuzzy belongingness, not sharp membership. All concepts of fuzzy set theory apply here, as learning is too complex to admit strict definitions in terms of necessary and sufficient conditions.

I begin in the first section with a historical conceptual analysis where I compare various definitions of “learning” in the specialized literature, to identify which ones favor an affirmative answer and which a negative answer, and why. In the second section, I extract from this analysis some elements for a broader definition that provides a conceptual foundation for a negative answer to Q. I end with some concluding remarks about some possible criticisms and ways to address them. What follows is just one possible analysis. Others could lead to different outcomes. Therefore, everything I say throughout the article is suggestive, preliminary, and tentative, open for debate under alternative analyses.

Fig. 1 Fuzzy partition (dashed curves) of the fuzzy set of learners into four families: Organisms with brains (mostly animals, vertebrates as well as invertebrates, especially insects), spinal vertebrates, some aneural organisms (paramecia, bacteria, and plants), and inorganic machines. The fuzziness of these families makes fuzzy set theory applicable to all of them, the family of learners included

Some Definitions of “Learning”

I begin with textual evidence of my claim that discussions over the definition of “learning” are not mere semantic discussions over the meaning of the term, but have a strong ontological import about the nature of learning. For example, Domjan (2015) wrote, “A universally accepted definition of learning does not exist” (p. 13). Haselgrove (2016) discusses various definitions of “learning” in the first chapter of his book, “What is Learning?” where he gives a similar diagnosis: “. . . there is no generally accepted definition of learning” (p. 2). Insofar as the term “accepted” refers to accepting a statement as true, these quotations seem to treat definitions of “learning” as views about the nature of learning.

Haselgrove (2016) also makes this diagnosis: “it turns out to be relatively difficult to pin down an entirely satisfactory definition of the phenomenon” (p. 1). Catania (2012) made a similar diagnosis: “From the start, we must face the fact that we won’t be able to define learning. There are no satisfactory definitions. But we won’t let that stop us” (p. 2). I won’t let it stop me either, but I take it that a satisfactory definition of “learning” is one that correctly expresses the nature of learning.

A debate in learning research that further supports my claim is found in Grau and Joynes’s (2005; see also Grau, 2014) frustration at years of being questioned about whether they “were examining *true* learning,” where some reviewers provided “demerits for falling outside traditional categories” (p. 46). Such questioning suggests a position on the nature of learning that certain behavioral changes in spinal vertebrates presumably do not satisfy. Grau and Joynes were thus criticized for displaying “conceptual confusion” (Machado, 2005, p. 28), making “a fundamental conceptual error” (p. 30), and not having “defined learning itself” (Reilly & Schachtman, 2005, p. 36). These criticisms are symptomatic of positions where such behavioral changes do not qualify as “*true* learning,” referring to the nature of learning as specified by a definition of “learning.” Grau and Joynes put it thus: “Frustrated by this continued assault, we have turned the question around and asked why learning should be so narrowly defined?” (p. 46). This rhetorical question suggests a debate over the definition of “learning” qua statement about the nature of learning.

In sum, discussions over the definition of “learning” are not mere semantic quibbles over the meaning of “learning” but ontological debates over the nature of learning. Two families of definitions of “learning” in these discussions are readily identifiable that denote deep disagreements over such nature: behavioral and nonbehavioral. In behavioral definitions, *at least* some behavior or performance change is defining “learning.” There may or may not be other defining properties. Thus, a behavioral definition may or may not be purely behavioral. In nonbehavioral definitions, in contrast, behavior does not define “learning” at all. In these definitions, “learning” is defined *only* by nonbehavioral properties.

I now discuss some members of each family in turn as orderly as I can, whether they favor an affirmative or negative answer to Q, whether they include IML, and why. My present purposes will make me partial to the latter. I am ultimately interested in definitions that include IML but are also plausibly applicable to animal learning. The list of definitions will be far from exhaustive, but I hope it will be representative.

Some Behavioral Definitions

An early example is Thorpe's (1943) definition of "learning" as "that process within the organism which produces adaptive change in individual behaviour as a result of experience" (p. 220). This definition qualifies as behavioral according to my criterion above if it defines "learning" as a kind of behavior change, even if produced by an internal process. It is unclear whether learning in this definition is *just* "the process within the organism," or also the behavioral change this process produces. If the latter, the definition qualifies as behavioral. Of course, it is not *purely* behavioral, as it appeals to causal internal processes, but it is behavioral nonetheless, as it includes behavior as defining of "learning," or so it seems.

Many psychologists would point out that not all learning is adaptive, if this term means an increase in reproductive success. Maladaptive behavior or misbehavior is quite common (e.g., addictions, anger conversion, attention-seeking behavior, reinforcement-delaying adjunctive behaviors), and much of it is learned, but it is far from obvious how it qualifies as "adaptive" in that sense of the term. Perhaps we should abandon this requirement if a definition of "learning" is to include maladaptive behavior. Of course, the term "adaptive" is fine if it means just "adjusting," "changing," or "flexible," but then these terms would be more precise.

The above definition seems to be sufficiently neutral about the nature of the internal process that (partly) defines "learning" to capture IML and, hence, favor a negative answer to Q. However, two features of the definition warrant a different interpretation. One is talk of "organism." Technically, in biology, this term refers to all living organisms, animal as well as nonanimal (to include bacteria, paramecia, and plants). If this term is used strictly in this sense, the above definition favors a negative answer to Q (insofar as nonanimal organisms do not have brains, or even nervous systems), but still excludes IML. If the term "organism" is used informally, as short for "whole animal," this definition warrants an affirmative answer to Q and, hence, conceptually excludes IML.

The other feature is Thorpe's talk of "individual behavior." Philosophically, the term "individual" is very broad: My chair, my table, and an atom are individuals in this philosophical sense. If the term is used in this way, the definition conceptually includes IML (and, hence, favors a negative answer to Q). However, it is customary among biologists and psychologists to use "individual" as synonymous with "living organism" (in some cases, "animal," and yet in many cases "human"), which would exclude IML.

However, the author views habituation as "the simplest type of change in behaviour which can be regarded as evidence of learning" (Thorpe, 1943, p. 220), and defines "habituation" as follows: "an activity of the central nervous system whereby innate responses to mild shock and warning stimuli wane as the stimuli continue for a long period without unfavourable results" (p. 221). If habituation is a representative type of learning fundamentally similar to others, this definition suggests a conception of learning strongly tied to beings with a central nervous system. If this interpretation is correct, this author's definition of "learning" favors an affirmative answer to Q and, hence, excludes IML (cf. Marsland, 2009). Unfortunately, the author did not give a definition of "associative learning" in that article, but it would be odd if he conceived it in a way that excluded nervous systems. All in all, then, this definition seems to be strongly tied to nervous systems.

The classical, most often used behavioral definition of “learning” is Kimble’s (1961): “. . . learning refers to a more or less permanent change in behavior which occurs as a result of practice” (p. 2). This definition does not specify what kind of behavior whose change constitutes learning, but the author’s exclusive focus on *animal* learning is evident throughout his popular textbook, especially in the section entitled “The Adaptive Character of Behavior” (pp. 18–19). The title indicates a restriction of behavior to adaptive behavior, which excludes maladaptive behavior. The author’s effort to make substantive contact with the Darwinian theory of evolution by natural selection is laudable, but evolution by natural selection is as much about maladaptive as it is about adaptive traits. All this supports my proposal to abandon the requirement of adaptiveness from a definition of “learning.” This issue aside, if that focus on animals is a valid indicator of a restriction of the concept of “learning” to whole animals, the same conclusion as before follows: The definition favors an affirmative answer to Q and, hence, excludes IML.

Lachman (1997) has pointed out that most textbook definitions of “learning” are like this classical one. For instance, Mazur (2017) gave this definition:

Although even specialists have difficulty defining the term *learning* precisely, most would agree that it is a process of change that occurs as a result of an individual’s experience. Psychologists who study learning are interested in this process wherever it occurs—in adults, school children, other mammals, reptiles, and even insects . . . psychologists study not only the *process* of learning but also the *product* of learning—the long-term changes in one’s behavior that result from a learning experience. (p. 2)

At first, this definition mentions “a process of change” without specifying a change in what. Only after distinguishing between process and product does the author specify a change in behavior as defining of “learning,” in which case the definition is behavioral. It is unclear, however, whether “learning” in this definition refers to just the process or also the product.

The behavioral character of these definitions does not preclude inclusion of other properties as defining of “learning” such as neurological processes. Mazur (2017) includes a discussion about brain and behavior, but it is unclear whether he views nervous system functioning as defining of “learning.” If he does, then his definition too prompts an affirmative answer to Q, to the exclusion of IML. Equally unclear is whether this author would view IML as true learning. His expression “and even insects” in the above quotation suggests that insect learning is far enough.

What Is Behavior?

Whether behavioral definitions of “learning” prompt an affirmative or negative answer to Q, and, if the latter, whether they include IML, hinges crucially on the definition of “behavior.” Trying to define “behavior” is at least as daunting as trying to define “learning,” if not more. In a survey, Levitis, Lidicker, and Freund (2009) found wide disagreements over the definition of “behavior” among behavioral biologists. The authors then offer their own definition, which they claim is an improvement over others: “the internally coordinated responses (actions or inactions) of whole living

organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes” (p. 108).

The relevant aspect of this definition of “behavior” for my present purposes is its restriction to *whole living organisms* (equivalently referred to as “individuals”), which clearly excludes inorganic machines. This restriction does not necessarily mean a restriction to animals, of course, unless, again, “living organism” is used as equivalent to “animal,” which I seriously doubt is the case of biologists. A behavioral definition of “learning” that uses this definition of “behavior” may therefore prompt a negative answer to Q (and hence include learning in spinal vertebrates and certain nonanimal organisms), but it surely excludes IML.

De Houwer, Barnes-Holmes, and Moors (2013) gave this definition of “learning”: “. . . learning can be defined as *changes in the behavior of an organism that are the result of regularities in the environment of that organism*” (p. 633; emphasis in original). In and of itself, this definition seems to be purely behavioral, but later the authors give this definition of “behavior”:

. . . we define the term *behavior* very broadly. It encompasses every observable response that a living organism can make, regardless of whether the response is produced by the somatic nervous system (e.g., pressing a lever), the autonomic nervous system (e.g., salivation), or neural processes (e.g., electrical activity in the brain). (p. 633)

This definition of “behavior,” as broad as its authors intended it to be, also restricts their definition of “behavior” and, to this extent, of “learning,” to living organisms, which excludes IML. This definition also prompts an affirmative answer to Q *if* by “organism” the authors mean “whole animal.” This definition would exclude learning in spinal vertebrates (much to Grau’s frustration) insofar as they are not whole animals. For the same reason, the definition would also exclude learning in nonanimal organisms such as bacteria, paramecia, and plants. But then again, such exclusions hinge on whether the term “organism” is restricted to whole animals. If the term encompasses nonanimal organisms, the definition would favor a negative answer to Q but would still exclude IML.

Another example is Pear (2016), who defines *learning* as “a dependency of current behavior on the environment as a function of a prior sensory-motor interaction with the environment” (p. 11), and *behavior* as “any neurological activity that is typically measured as motor (i.e., muscular or glandular) activity” (p. 11). This definition of “behavior” favors an affirmative answer to Q *if* by “neurological activity” the author means “organic nervous system activity” or “brain activity.” If not, the definition warrants a negative answer, but still excludes IML, as well as nonanimal organisms, insofar as organic neurological activity (thus conceived) is a property of animals.

A similar outcome obtains for other behavioral definitions of “learning” accompanied by a definition of “behavior.” Colman (2015) defined *learning* as “Any lasting change in behaviour resulting from experience, especially conditioning” (p. 416), but then defines “behaviour” as

The physical activity of an organism, including overt bodily movements and internal glandular and other physiological processes, constituting the sum total of

the organism's physical responses to its environment. The term also denotes the specify physical responses of an organism to particular *stimuli or classes of stimuli. (p. 83)

This author defines “organism” as “Any living thing, including a human or other animal, a plant, a fungus, a protist, a bacterium, or a virus” (p. 529). By explicitly including nonanimal organisms, this definition favors a negative answer to Q that includes paramecia, bacteria, plants, but still excludes IML, as the definition is restricted to living organisms. The definition, however, excludes spinal vertebrates if by “organism” the author means “whole organism.”

Kearney (2015) speaks of “behavior-changing experiences we call *learning*” (p. 30) and gives this behavioral definition of “learning”: “... any relatively permanent change in behavior that results from interaction with the environment” (p. 31). This definition seems to be purely behavioral in that it mentions no defining property of learning other than an *overt* behavioral change. A few pages earlier, however, the author wrote this:

But we also have internal, covert, private behaviors. These internal, covert behaviors include physiological acts of our bodies, such as the beating of our hearts and the digestion going on in our stomachs. Even emitting brainwaves are internal behaviors that are not so easily observed. Usually medical instruments of some kind are needed to observe and measure covert physiological behaviors, but just because one sees the actions of these internal bodily organs directly, this doesn't mean that they are not behaviors. (p. 28)

It seems clear that, according to this author, at least some behaviors are brainwaves, which implies that at least they require at least a brain. It remains unclear whether “brainwaves” are included in the author's definition of “learning” above, however. If it is, this definition of “learning” might favor an affirmative answer to Q. Equally unclear is whether brainwaves are necessary for learning other behaviors that are not brainwaves.

The same author also gives this definition of “behavior”: “behavior is any external or internal observable and measurable act of an organism” (Kearney, 2015, p. 27). The author does not clarify what he means by “organism,” but if he follows standard practice and means “living organism,” his definition excludes IML, although it might include spinal vertebrates and nonanimal organisms, if his use of the term “organism” is not restricted to whole animals. In addition, the notion of “covert” or “internal” behavior is too vague to be of any use, so I shall dispense with it and view of all behavior as overt.

A similar outcome obtains with behavioral definitions of “learning” that seem to be less purely behavioral. Pierce and Cheney (2017) gave this definition: “*Learning* refers to the acquisition, maintenance, and change of an organism's behavior as a result of lifetime events. The *behavior* of an organism is everything it does, including private and covert actions like thinking and feeling” (p. 1). A kind of performance change seems to be defining of “learning” in this definition, which makes it behavioral according to my criterion. Moreover, the definition of “learning” and the definition of “behavior” are both restricted to at least organisms, which are usually conceived as living things. Both definitions thus exclude IML, although they might include

nonanimal organisms and, hence, favor a negative answer to Q, but if by “organism” the authors mean “whole organism,” their definition would exclude learning in spinal vertebrates.

The authors also wrote this: “Learning also involves *neuroplasticity*—alterations in the brain that accompany behavior change and participate in the regulation of behavior” (Pierce & Cheney, 2017). It is unclear whether by “involves” these authors mean “defined by.” If yes, their definition would still qualify as behavioral, although not purely behavioral. This view is entirely compatible with the basic tenets of behavior analysis, as long as neuroplasticity (or any nervous system process) is not viewed as playing any causal-explanatory role in behavior. In any case, like previous behavioral definitions, this one seems to entail an affirmative answer to Q, and thus the exclusion of IML, insofar as learning is restricted at least to living organisms, as only they are capable of *organic* neuroplasticity.

Thus, much hinges on whether definitions of “behavior” and, to this extent, learning behaviorally defined, are restricted to living organisms. There are some such restrictions. For example, Millikan (1993) proposed the following:

A behavior is, I suggest, at least the following: 1. It is an external change or activity exhibited by an organism or external part of an organism. 2. It has a function in the biological sense. 3. This function is or would be normally fulfilled via mediation of the environment or via resulting alterations in the organism’s relation to the environment. . . . Requirement 2 is the central one. (p. 137)

Her restriction of the concept of behavior to living organisms is quite explicit. Therefore, any behavioral concept of learning based on that concept of behavior will at least and most clearly exclude IML. It will also exclude nonanimal behavior if she restricts living organisms to animals. It may well include spinal vertebrates, insofar as they are living, but they are not whole. If she restricts behavior to whole organisms, any definition of “learning” based on her concept of behavior would also exclude learning in spinal vertebrates and inorganic machines.

In their survey, Levitis et al. (2009), p. 110) reported approval ratings for various examples and features of behavior among members of different societies of behavioral biologists. Ratings interpreted as strongly approving clearly tended to restrict behavior to whole living organisms. Only members of the International Society for Applied Ethology strongly approved restricting behavior to animals. Far more widely and strongly approved was the notion that behavior was a feature of individuals rather than their parts, which seems to mean a restriction of behavior to whole living organisms. All the examples and features of behavior in the survey comprised living organisms. There was no mention of inorganic machines anywhere (although the authors once spoke of “an animal’s machinery” [p. 103], indicating the possibility of organic machines, which warrants my insistence in speaking specifically of *inorganic* machines).

Likewise, Baum (2013) formulated this basic principle for a definition of “behavior,” among others: “(a) Only whole living organisms behave” (p. 285). This principle explicitly restricts a definition of “behavior” to whole living organisms. A behavioral definition of “learning” supplemented by a definition of “behavior” based on this principle thus favors a negative answer to Q (if “living organism” is not used as

synonymous with “animal”) that includes nonanimal organisms, but excludes spinal vertebrates and IML.

Dretske (1988) gives a definition of “behavior” that is not restricted to living organisms and, hence, is more suitable for my present purposes. He distinguishes a behavior of, from what happens to, a system, without restricting behavior to living organisms:

Animal behavior is what animals do. Human behavior is what humans do. If plants and machines do things, then whatever they do is plant and machine behavior. . . . When a rat moves its paw, that is something the rat does, a piece of rat behavior. When *I* move its paw, the paw still moves, but the rat doesn’t move it. There is no rat behavior. Indeed, I could be moving the paw of a dead rat and dead rats do not behave. . . . The suggestion is that behavior is endogenously produced movement, movement that has its causal origin *within* the system. . . . (pp. 1–2)

More briefly: “Behavior, then, is to be identified with a complex causal process, a structure wherein certain internal conditions or events (*C*) produce certain external movements or changes (*M*)” (Dretske, 1988, p. 21). This definition, of course, does not exclude the possibility that *C* in turn is caused by some external environmental conditions *E* (see pp. 22–27), which makes it applicable to learned behavior (the definition applies to nonlearned behavior as well).

This concept of behavior can be used to give a behavioral definition of “learning” that is not restricted to living organisms and, hence, includes IML: Learning is a process of modification of a system’s $I \rightarrow M$ causal relations (“ $\rightarrow$ ” denotes “causes”) by changes in *S*’s *E*. I replaced *C* for *I* to denote internal states. This definition would seem to be promising for my present purposes. Unfortunately, it faces an unresolved issue I mention in the next section that puts behavioral definitions in danger of being invalid.

Summing Up

The conceptual pattern that seems to emerge from the preceding analysis is that behavioral definitions of “learning” favor an affirmative answer to Q if restricted to *whole animals*. It is unclear whether authors use “organism” as synonymous with “whole animal.” If they do, their behavioral definitions of “learning” would favor an affirmative answer to Q. If by “living organisms” they mean to include nonanimal organisms (paramecia, bacteria, and plants), their definitions of “learning” favor a negative answer to Q. They would still exclude learning in spinal vertebrates, *if* the emphasis is on *whole* organisms, as well as inorganic machines.

Table 1 summarizes this analysis by showing the conceptual restrictions I have identified in the quoted behavioral definitions of “learning” when supplemented by the quoted definitions of “behavior” (columns), whether they favor an affirmative (A) or negative (N) answer to Q, and whether they include (✓ = yes, ✗ = no) spinal vertebrates (SV) or inorganic machines (IM). The conceptual restrictions can be to whole animals, animals (whether or not whole), whole living organisms, and living organisms (whether or not whole).

Table 1 Summary of the conceptual analysis of behavioral definitions of “learning.” Columns: Possible conceptual restrictions of the resulting definitions (whole animals, animals whether whole or partial including missing key anatomical parts of their nervous system, whole living organisms, and living organisms whether or not partial). A: Affirmative answer to Q. N: Negative answer to Q. SV: Spinal vertebrates. IM: Inorganic machines

			Conceptual restrictions			
			Whole animals	Animals	Whole living organisms	Living organisms
Answer to Q	A	✓				
	N			✓	✓	✓
Includes	SV	✗		✓	✗	✓
	IM	✗		✗	✗	✗

Only a restriction to whole animals (first column, left to right) favors an affirmative answer to Q and, hence, excludes spinal vertebrates, nonanimal organisms, and inorganic machines. A restriction to animals (second column), without specifying whether they are whole, logically allows for the possibility of partial animals, which warrants a negative answer to Q. Thus, this restriction includes spinal vertebrates, but neither nonanimal organisms nor IML. A restriction to whole living organisms (third column) favors a negative answer, and includes nonanimal organisms but neither spinal vertebrates nor IML. A restriction to living organisms, whether or not whole (fourth column) favors a negative answer to Q and includes spinal vertebrates, but excludes IML. Whether a living organism is whole or partial, then, makes no difference in how Q is answered, but makes a difference in whether spinal vertebrates are included.

Some Nonbehavioral Definitions

Despite the dominance of behavioral definitions of “learning” in textbooks, many psychologists reject such definitions in favor of nonbehavioral definitions that appeal *only* to certain internal processes, regardless of whether or not they cause any performance. In terms of the symbolism above, nonbehavioral definitions propound only the causal relations $E \rightarrow I$ as defining of “learning,” regardless of whether $I \rightarrow M$ relations also obtain. Changes in M are just an empirical indicator of changes in I , but do not define “learning.”¹

¹ A thoughtful reviewer asked, “How can one observe learning where there is no change in behavior?” Nonbehavioral definitions allow for this answer: Observe the internal processes in behaving as well as nonbehaving subjects; if comparable, the latter can be said to have learned. The research of Eric Kandel and colleagues (e.g., Carew, Walters, & Kandel, 1981) suggests how it could be done. Train various specimens of *Aplysia*, immobilize some, and observe the internal processes (postsynaptic excitatory potentials on the motor neurons that mediate the response, transmitter release from sensory neurons, and concentration of intracellular calcium and second messengers) in all the specimens. The immobilized ones can be said to have learned if such processes are comparable to those observed in the behaving subjects. Moore’s theorem (also mentioned by the reviewer) becomes inapplicable, as it is restricted to black boxes, whereas organisms whose internal processes have been observed are not black boxes. There is no need to include IML anywhere here. As clarified in the introduction, I intend the view that IML is true learning as a conjecture in its own right, regardless of whether and how it benefits the study of animal learning, although I do not exclude the possibility of some benefit either.

The rationale behind this kind of definition is what Dickinson (1980) called metaphorically “the problem of behavioural silence” (see pp. 15–18). If behavioral change defines learning, *whether or not partly*, several phenomena widely regarded as involving learning (e.g., latent inhibition, sensory preconditioning, second-order conditioning in pigeons with a tone as a conditioned reinforcer, US preexposure effect, and learning in curarized animals [e.g., Cousins, Zamble, Tait, & Suboski, 1971]) do not qualify learning. This implication, the rationale concludes, is counterintuitive. There is no obvious refutation of this rationale, which gives behavioral definitions trouble. As Burgos and Killeen (2018) have argued, the behavior analysts’ efforts to refute this rationale for mentalistic (dualism, unobservability, and impracticality) are ultimately ineffective.

Lachman (1997) has criticized behavioral definitions of “learning” as “unsatisfactory,” based on that rationale: “. . . learning may not include a change in behavior” (pp. 477–478). Clearly, if learning may not include a change in behavior, learning can occur without behavior. Hence, behavior does not define “learning.” On this basis, he proposes the following definition as “new”: “. . . learning is the process by which a relatively stable modification in stimulus-response relations is developed as a consequence of functional environmental interaction via the senses” (p. 479). It is unclear, however, how this is not a behavioral definition, if stimulus-response relations are behavioral in nature. This author also views this definition as “Applicable to every level of living organism,” which seems as restricted to living organisms as behavioral definitions are.

In the chapter “What is Learning?” in their book, Olson and Hergenhahn (2016) began with a modified version of Kimble’s (1961) classical behavioral definition of “learning,” but then gave a clearly nonbehavioral definition in a section entitled “Must learning result in a behavioral change?” where they clearly accept the behavioral-silence rationale:

. . . whatever we study in psychology, must be expressed through overt or covert behavior, but this does not mean that the behavior we study *is* learning. We study behavior so that we can make *inferences* concerning the process believed to be the cause of the behavioral changes we observe. In this case, that process is learning. Most learning theorists discussed in this book agree that the learning process cannot be studied directly; instead, its nature can only be inferred from changes in behavior. . . . most learning psychologists look on learning as a process that mediates behavior. For them, learning occurs as the result of certain experiences and precedes changes in behavior. In such a definition, learning is given the status of an *intervening variable*. . . a theoretical process that is assumed to take place between the observed stimuli and responses. (pp. 2–3; emphases in original)

This definition of “learning” is clearly nonbehavioral according to my criterion above in that behavior does not define (is inessential to), even if empirically indicative of, learning (do not confuse “learning is . . .” with “learning is evidenced by . . .”). Learning according to nonbehavioral definitions is *just* certain covert, normally publicly unobservable, internal processes that play a causal role when evidenced in behavior, which is not always the case. Sometimes, according to this definition,

learning is not expressed in behavior. Therefore, behavior is not defining of “learning” (i.e., inessential to learning).

Nonbehavioral definitions of “learning” favor an affirmative answer to Q (and, hence, excludes IML) only if intervening variables are restricted to the functioning of a nervous system. If intervening variables are interpreted neurobiologically, they may favor a negative answer insofar as this interpretation includes learning with an incomplete nervous system (e.g., without certain neuronal structures). Here are some examples.

Some Neurobiological Definitions

It is important that Olson and Hergenhahn (2016) failed to mention that Kimble (1961) referred to his behavioral definition as “factual” in that it defined “learning” by “observable events in the physical world” (p. 2). These definitions, he claimed, “make no effort to describe underlying mechanisms or to identify the “true nature” of learning” (p. 8). He referred to definitions of learning that “attempt to accomplish such objectives” (p. 9) as “theoretical,” without repudiating them. He clearly viewed the “true nature” of learning to be given by underlying mechanisms, not behavior change.

Kimble (1961) further stated that “[m]any of the theoretical definitions of learning are physiological” (p. 9). For example, he cited the following definition: “Learning is the process of the formation of relatively permanent neural circuits. . . (Bugelski, 1956, p. 120).” This definition is reminiscent of Thorndike’s (1913) connectionist definition: “Learning is connecting. . . . There are millions” of connections, including “connections leading to actual conduction in neurones” (p. 54; “millions,” of course, vastly underestimates the presently estimated number of connections in an average human brain); “The process of learning is one of simple making and keeping connections and readinesses to conduct” (p. 56).²

Also reminiscent of such connectionist character is Hull’s (1943) definition, given in the section “The Problem and General Nature of Learning”: “The essential nature of the learning process may . . . be stated quite simply . . . the process of learning consists in the strengthening of certain . . . connections as contrasted with others, or in the setting up of quite new connections” (pp. 68–69).

The connectionist character of these definitions means that they propound a strong link between learning and nervous systems, which suggests an affirmative answer to Q and, to this extent, an exclusion of spinal vertebrates, nonanimal organisms, and IML. Definitions that postulate this link have continued to appear decades later. Dudai (1989) defines “learning” as “an experience-dependent generation of enduring internal representations, and or experience-dependent lasting modification in such representations,” but clarifies that “representation is an inherent and fundamental function of all nervous systems” (p. 6). This definition views a

² These last two definitions, and others quoted earlier (Thorpe’s) and later (Bolles’s), conceive learning as a process, predating Lachman’s (1997) alleged “new” definition by several decades. Here are others: “. . . learning is a process in the behavior of the individual” (Skinner, 1950, pp. 195–196); “Learning is a process by which an organism benefits from experience so that its future behavior is better adapted to its environment” (Rescorla, 1988, p. 329); “. . . learning as a process of accumulation where incorrect response tendencies remain constant and correct response tendencies increase with practice” (Mazur & Hastie, 1978, p. 1256). In general, it seems plausible to conceive learning as a process, and I will adopt this view later on.

certain nervous system function (representing the external environment) as defining of learning, which entails an affirmative answer to Q and, hence, excludes IML. The definition does not mention any kind of behavior change as essential to learning, for which I take it as nonbehavioral, although the definition does not deny the possibility of nonneural representations.

Menzel, Greggers, and Hammer (1993) gave this definition: “Learning is a property of the nervous system in which informational status of stimuli is changed as a consequence of being passively or actively exposed to stimuli and their combinations” (p. 80). Likewise, Higgins and George (2013) submit the following definition: “*Learning* is defined as new information acquired by the nervous system and observed through behavioral changes” (p. 217).

Dukas (2009) is more concrete in specifying neuronal process as defining of “learning”: “Learning may be defined as the acquisition and retention of neuronal representations of new information” (p. 7; see also Dukas, 2018, p. 257). Although this definition does not specify what the neuronal includes (a single neuron is neuronal), it is common to identify neuronal representations with quasi-stable patterns of synaptic efficacies in relatively large neuronal circuits. Nor does it specify behavior as essential to learning.

Likewise, Domjan (2015) gives the following definition: “Learning is an enduring change in the mechanisms of behavior involving specific stimuli and/or responses that results from prior experience with those of similar stimuli and responses” (p. 13). He then clarifies that “. . . the preceding definition attributes learning to a change in the *mechanisms of behavior*, not to a change in behavior directly. . . Learning is defined in terms of a change in the mechanisms of behavior to emphasize the distinction between learning and performance” (p. 13; emphasis in original). The nonbehavioral character of this definition is clear: Behavior is indicative of the occurrence, not defining, of learning.

This definition, like the other nonbehavioral ones, makes a distinction between learning and performance where the latter does not define “learning.” Exactly what are the mechanisms that define “learning” remains under investigation, but nervous-system mechanisms are primary candidates, as Domjan suggests in his Figure 1.5 (p. 16), where he depicts neural systems, circuits, cells, and synapses, and constituting “Levels of analysis of learning.”

In general, these nonbehavioral definitions are specific about the nature of the covert mechanisms of behavior that define “learning”: They are neurobiological. They favor an affirmative answer to Q, if the term “neurobiological” is used as shorthand for “nervous system” in the flexible sense I am using this expression here, to refer to at least some major anatomical part of a brain. If these definitions define “learning” as the functioning of one or more major brain circuits, they exclude spinal vertebrates (as their brains are effectively disconnected from the spinal cords), nonanimal organisms (as they have no nervous system), and IML (as nervous systems are organic).

Some Nonneurological Definitions

Other nonbehavioral definitions are more abstract about the nature of the underlying mechanisms of behavior that define “learning.” To this extent, they favor a negative answer to Q, although they do not necessarily include IML. For example, Sahley, Rudy,

and Gelperin (1984) propose that “Associative learning is a theoretical construct, an inference one makes on the basis of an observed relationship between an organism’s behavior and its known past experience” (p. 243). This definition does not restrict learning to a certain kind of brain process, which favors a negative answer to Q, but it still seems to be restricted to living organisms, which excludes IML.

Bolles (1976) also gives another clearly nonbehavioral definition that is more abstract regarding the nature of the internal processes postulated to be essential to learning: “. . . learning may be defined as the process of associating ordered events. Note that there is no evidence of learning at the time learning occurs; evidence of learning consists of a subsequent change in behavior” (p. 21). In this definition, a behavioral change only gives evidence of learning, but does not define “learning.” The definition per se does not specify whether it is restricted to living organisms, but its immediately preceding context makes the author’s focus on animal learning clear:

If an experimental situation provides an ordering of events such that one event follows another in a lawful manner, and if an animal’s behavior changes as a result of this arbitrary ordering, then we may say learning has occurred. The learning itself is difficult to identify because it is not an observable thing or relationship but a hypothetical process, which has been variously conceived as forming a neural connection, building a habit, establishing an association, or acquiring an expectancy. But however the process is conceived, the change in behavior merely constitutes evidence of learning. (p. 21).

Such a focus on animal learning favors an affirmative answer to Q and, therefore, an exclusion of IML, or so it seems. Bolles gave much importance to biological evolutionary considerations in learning, which strongly suggests a restriction of the several conceptions of learning he mentions to living organisms. He would have likely been perplexed if asked whether inorganic things, perhaps even nonanimal organisms, were capable of building of habits, establishing associations, or acquiring expectancies (of course, they are incapable of forming *organic* neural connections). The definition might also include spinal vertebrates if brains are not necessary for those things.

This definition raises the additional difficulty that it is unclear whether establishing associations, forming memories, acquiring expectancies, or being surprised require a whole nervous system. I still have not found a psychologist of learning who defines “learning” nonbehaviorally clarify this. Perhaps, if pressured, they might admit that ultimately all of that requires a nervous system, in which case these definitions of “learning” favor an affirmative answer to Q and, hence, exclude IML. But if these things only require a brain, then this definition would favor a negative answer.

Another possibility is to adopt a computationalist view about learning, according to which the internal processes that define “learning” are computational in nature. In this case, such definitions would favor a negative answer to Q and include IML. This possibility is expressed in Gallistel’s (1990) claim:

Learning is intimately connected to computational machinery that extracts information with a particular formal structure from particular sensory inputs,

independent of the immediate utility the information may have and independent of the uses to which it may subsequently be put by diverse readout mechanisms. (p. 88)

It is unclear whether the author meant this claim as a definition of “learning,” but if he did, it expresses a view of learning as computational in nature, which favors a negative answer to Q, insofar as the computations that constitute learning are physically realizable by systems other than nervous systems, at least in principle. Likewise, in a “The Nature of Learning,” Gallistel and King (2010) identify “two different stories about what learning basically is”:

In the first story about the nature of learning, which is by far the more popular, one, particularly in neurobiologically oriented circles, learning is the rewiring by experience of a plastic brain. . . . In the second story, learning is the extraction from experience of information about the world. . . . In the first story, the brain has the functional architecture of a neural network. In the second story, it has the functional architecture of a Turing machine. (p. 187)

These authors prefer the second story, where learning is viewed as digital symbolic Turing-machine computation physically realizable in different ways, nervous system being one of them. This approach is a form of computationalism, the thesis that something (e.g., learning qua internal process) is computational in nature. This approach favors a negative answer to Q and includes IML, but this does not necessarily make it suitable for my present purposes, as I clarify in the next section.

In the first story, neoconnectionists propound neural-network modeling as the best way to theorize about cognition and define “learning” in a way inspired by the structure and functioning of brains. They usually define “learning” as changes in the connection weights or “free parameters” of a neural network. Here is one definition: “Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded” (Teuscher, 2002, p. 7). Here is another: “Setting the connection strengths correctly is what network learning is about” (Anderson, 1995, p. 55). Mangel (1993), in a chapter of a book on insect learning, modifies Dudai’s definition quoted above to express it in neoconnectionist terms:

In the language of neural networks . . . an “enduring internal representation” is a description of the external world based on connections between different neuronal groups and rules for modifying those connections. Learning represents changes in the pattern of connections or the rules for modifying those connections. (pp. 158–159)

These definitions make no distinction between natural and artificial neural networks, but I could use it here: The former are organic, the latter are not and, hence, not only favor a negative answer to Q, but also include a large class of IML (deep learning in artificial neural networks). If these authors do not see this difference as fundamental, as view natural neural networks as fundamentally

identical to artificial neural networks, then they are committed to another form of computationalism (connectionist), which is equally unsuitable for my present purposes, as I clarify in the next section.

Similar considerations apply to Enquist and Ghirlanda's (2005) assertion that "A tentative definition of *learning* might be 'changes in behavior caused by experience,'" but:

. . . a better definition of learning is "changes in memory caused by experience," whether or not such changes are immediately reflected at the level of behavior. . . . In neural network models, memory and learning can be given precise interpretations reflecting this definition. Memory is equal to connection weights, the analog of synapses in nervous systems . . . learning amounts to changes in the weights. (p. 130)

What does make the second definition better than the first? The authors give the behavioral-silence rationale mentioned above for non-behavioral definitions of "learning": "An experience, however, may change later behavior without this being immediately apparent" (p. 130). This definition of learning also favors a negative answer to Q and includes a major class of IML.

However, there are nonneural forms of IML that neoconnectionist definitions of "learning" exclude. One widely studied example is genetic algorithms, to which Mitchell (1997) dedicates a whole chapter (pp. 249–273), and were first proposed by Holland (1975). In brief, genetic algorithms are search algorithms inspired by evolution by natural selection with genetic inheritance, where genotypes are simulated as bit strings that encode certain magnitudes simulating certain quantitative phenotypic traits and are evaluated through a fitness function. The fittest individuals are more likely to be selected for bit-string recombination and replication in a next generational cycle. It is standard practice to distinguish genetic algorithms from artificial neural networks, but to refer to the former as learning systems. Therefore, artificial neural network definitions of "learning" in neoconnectionism exclude an important class of IML. I therefore need a broader definition.

In his landmark book about machine learning, Mitchell (1997) defined "learning" as follows:

For the purposes of this book we will define learning broadly, to include any computer program that improves its performance at some task through experience. Put more precisely,

Definition: A computer program is said to *learn* from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*. (p. 2)

This definition is widely used in the machine-learning literature and includes artificial neural networks as well as genetic algorithms (to which this author dedicates a full chapter). This definition obviously includes IML and, hence, favors a negative answer to Q. As an added bonus, the definition is not restricted to artificial neural networks

strictly conceived, but also includes kinds of IML that some contrast to artificial neural networks, such as genetic algorithms.

But then again, Mitchell's definition is not necessarily suitable for my present purposes. Psychologists of learning who define "learning" nonbehaviorally, based on the behavioral-silence rationale, would take issue with an improvement in performance as defining of "learning." Even psychologists who define "learning" behaviorally would also take issue with this requirement insofar as, again, many instances of learning do not involve improvement in any clear sense (e.g., substance abuse, maladaptive behaviors, misbehaviors). Another issue is that applying this definition to living organisms commits one to some form of computationalism, which to me is not ideal, as I clarify next.

Towards an Alternative Definition of "Learning"

I have given enough textual evidence to derive some elements for a conceptual foundation for a negative answer to Q that includes IML via an alternative definition of "learning." The pattern I have detected in the previous section is that behavioral definitions of learning as a kind of behavior change are unsuitable, not only because they tend to be restricted to biological organisms (mostly animals), but also because there still is no compelling refutation of the rationale behind nonbehavioral definitions, namely, the behavioral-silence argument. Neurobiological nonbehavioral definitions are unsuitable too, as they tend to restrict learning to organisms with nervous systems.

Consequently, only certain nonneurobiological nonbehavioral definitions of "learning" are more suitable for my present task. Mitchell (1997), as Gallistel (1990), and Gallistel and King (2010) definitions above, are clearly suitable for IML and, hence, favor a negative answer to Q. However, they are not suitable for living organisms without forcing a commitment to some form of computationalism, a theoretical hypothesis that holds that certain phenomena, such as the mind, learning, and its underlying brain functioning are computational in nature.

There are currently too dominant forms of computationalism, Turing-machine and connectionist. Unfortunately, they often are seen as mutually exclusive (e.g., Gallistel & King, 2010; cf. Piccinini & Scarantino, 2011), which is too divisive for the kind of integrative conceptual foundation I seek here. I therefore rather not take sides on this debate and exclude for now any reference to computations or processing (whether digital or analog, sequential or parallel), information extraction, and representation (whether centralized or distributed) from the elements of the conceptual foundation I seek. If the two forms of computationalism can ever be coherently integrated, the result can be added to the present analysis (nothing in it precludes either form of computationalism).

My defense of IML here does not imply any form of computationalism. One thing is to say that inorganic machines can learn, my claim here, but quite another that animal learning and its underlying brain functioning are computational in nature, wherein lies the aforementioned controversy. These are two quite different, logically independent assertions.

Processes

What is left? Plenty, as it turns out. One severely underdeveloped defining property of learning is its process character. Developing this feature is a good starting point towards the kind of conceptual foundation I seek. Definitions of “learning” as a process do not elaborate what a process is. Here is first step towards amending this situation:

A process is a coordinated group of changes . . . an organized family of occurrences that are systematically linked to one another. . . . A process consists in an integrated series of connected developments unfolding in conjoint coordination. . . . Processes are correlated with occurrences or events: Processes always involve various events, and events exist only in and through processes. Processes develop over time. (Rescher, 1996, p. 38)

Thus, as a process, learning, according to this definition of “process,” is a coordinated group of changes or organized family of events systematically linked to one another, a series of connected developments unfolding in conjoint coordination. Of course, this is exceedingly imprecise. One major detail is how processes get their organization, systematicity, or integration (i.e., what a process’ connectivity amounts to).

Probabilistic Causal Processes and Interactions

One possibility is *causation*, a topic with a wealth of conceptual resources teeming with analytic possibilities psychologists of learning have not tapped yet, but can help articulate a technically more advanced concept of learning as a process. For example, recent views of causation characterize it not as a relation between events, but as a process: “. . . the core idea of process theories of causation is that causation can be understood in terms of causal processes and interactions” (Dowe, 2009, p. 213; see also Salmon, 1984, 1998). The preferred analogy in these theories is not the causal “chain,” which suggests discontinuity, but the causal “rope,” which suggests *continuity* (e.g., a rolling billiard ball).³

In one of these theories (Dowe, 2009), “a causal process is a world line of an object that possesses a conserved quantity,” whereas “a causal interaction is an intersection of world lines that involves an exchange of a conserved quantity” (p. 219). Paradigmatic examples of conserved quantities are linear momentum, mass-energy, angular momentum, and electric charge. Paradigmatic examples of causal processes thus conceived are material particles in motion, light pulses, and sound waves. I would also count protein synthesis, natural selection, development, and learning as other examples. Process theories of causation have been combined with a probabilistic view of causation as probability-raising, into the most powerful (albeit still imperfect) account of causation currently available. On these hybrid views, causation is a spacetime intersection of causal processes (at least two) where one raises the probability of the other.

Learning (nonbehaviorally defined) can therefore be conceived as a probabilistic causal interaction of causal environmental and internal processes. But this still is

³ Such continuity does not necessarily skew the present proposal towards analog computation, as the continuity could be viewed as just an approximation implemented by binary computations.

imprecise. Myriad causal interactions (in that sense) in nature do not qualify as learning (e.g., various sorts of collision, emission and absorption of photons, epigenetically determined protein synthesis, single unconditioned reflexes) meet this definition. More precisely, learning is a *kind* of process, perhaps a set of processes. What features do such processes have? A key feature, I propose, is that they are describable in certain ways. Here is a partial list from the IML literature that apply to all the learners depicted in Fig. 1, without forcing a commitment to any form of computationalism (I use the symbolism introduced above with Dretske's (1988) definition of "behavior"):

- Linear or nonlinear (and within this category, monotonic or nonmonotonic).
- Stochastic (deterministic learning is a limit case where $p = 1.0$).
- Inductively biased (causal dependence on previous interactions with E , through some time-sensitive retention device R).
- Internally modifiable by E (i.e., I as causally dependent on E).
- Involving at least one learner, an individual physical system that "exhibit[s] enough structural complexity and internal articulation to make the internal-external difference reasonably clear and well motivated" (Dretske, 1988, p. 11).

Learners

The concept of a learner is another severely underdeveloped concept in the psychology of learning. In brief, in terms of the above characterization of causality as interaction between causal processes, a learner is the object in learning as a causal process. But more precision is needed. How much is "enough" in the above characterization is largely arbitrary. Here is one abstract, simple, preliminary, and tentative possibility (there could be many others), as a first general, noncircular notion of a learner that includes living organisms and inorganic machines. A *potential minimal learner* L is a physical system consisting of at least (see Fig. 2):

One detector D of changes in E that can be in at least two temporary states d_1 and d_2 (probabilistically caused by changes in E);

Fig. 2 Diagram of a minimal learner L as a system that could in principle learn. E : Environmental external process. D : L 's detector device of changes in E . d_1, d_2 : Possible temporary states of D . I : Internal part of L . i_1, i_2 : Possible temporary states of I . R : Retention device (in some cases, may be internal to L ; in others, it may not, which allows for extended cognition). $r_1, \dots, r_n$: Possible temporary states of R (more persistent than the states of D and I). Dashed arrows: Probabilistic causation. Solid arrows: Structural links that allow for causal probabilistic processes

A one-direction structural link from D to at least one internal part I of L (one-headed solid arrow) that can be in at least two temporary states i_1 and i_2 (probabilistically caused by changes in D);

A two-direction structural link between I and R (double arrow), a retention or “memory” component that can be in multiple temporary states ($r_1, \dots, r_n$) more persistent those of E , D , and I ; in some cases, R may be internal to L ; in others, it may be external, as in extended cognition (see Clark & Chalmers, 1998).

The last feature allows L to be inductively biased, by modulating changes in I by D . L as depicted here is a very rough approximation to the notion of a minimal learner, give or take a few parts and relations. It is very abstract, but that is the idea. I therefore left open all the specifics about all the components, their particular nature (organic or inorganic), and their causal relations, as well as the features of E . All these specifics are to be filled in for particular learners under particular environmental conditions, in ways that can suit any form of computationalism.

Thus, E might include laboratory experimental conditioning procedures, or conditions in L 's natural environment, or tasks given to a Turing machine, an artificial neural network, or a genetic algorithm. D includes animals' sensory receptors, whatever means paramecia, bacteria, and plants have to detect their local external environmental conditions, but also inputs to a Turing machine, an artificial neural network, or a genetic algorithm. I includes the various internal organs of animals, such as their nervous systems, as well as the internal molecular structure of a paramecium, bacterium, or plant, but also a Turing-machine table, the hidden layers of an artificial neural network, or the genotypes of a genetic algorithm. R includes what is usually called “memory” (if the reader needs to use this term; I do not, nor is it forced by my characterization of L). In animals, R would consist of a set of synaptic efficacies (according to the dominant view in neuroscience). In paramecia, bacteria, and plants, R would consist of other different kinds of various internal protein configurations. In artificial neural networks, R consists of the set of connection weights. In genetic algorithms, R is achieved by the fitness-dependent retention of previous bit strings through recombination. In a Turing machine, R corresponds to symbols written on a tape, and so on.

All the learners depicted in Fig. 1 meet this characterization. Of course, as depicted, L cannot perform overtly (“behave”) in any way, but then again, this is inessential to learning according to nonbehavioral definitions. L could receive an external performance device P that can be in at least two different temporary states, probabilistically caused by i_1 and i_2 , but then again, P would be inessential to leaning. It remains to be seen whether L is too inclusive and allows for systems that do not qualify as learners. For example, it remains unclear whether L thus conceived excludes expert systems that do not learn. But then again, this characterization of L is a rough first approximation. In any case, how do we know a learner has learned (see Note 1), if it cannot behave? By looking at is

Concluding Remarks

To conclude, I will address three possible criticisms to my present proposal. First, my above characterizations of a learner (*L*) above might be criticized for triggering a vicious circle. However, there is no circularity: I have not defined “learner” as “something that learns,” but as a system with certain structural and functional features abstracted from the learners depicted in Fig. 1. That is to say, being a learner, like being fragile, is a *dispositional* concept: Something can be a learner without ever learning, by having such features, just like something can be fragile without ever breaking, just by having a certain molecular structure. Therefore, no circularity follows from defining “learner” in this way. Perhaps “learner” is an unfortunate name for giving the impression of circularity, but I cannot think of a better one.

Second, my view of IML as true learning might be criticized for confusing a model with what is modeled. The rationale behind this criticism is that IML is a model of animal learning; therefore, it is not true, real learning. However, this rationale is fallacious. A scale model of a car is a real thing, but is it a car? Maybe, maybe not. It depends on the model. If it has no engine and cannot roll (its wheels are glued), it is not a car. But if it has an engine and rolls (e.g., by remote control), then it *is* a car, not of the same type people drive, to be sure, but a car nonetheless. Animal learning models of, say, drug addiction constitute real learning and drug addiction, even if obviously different from human learning and drug addiction, and are no less real for being models. The key feature of a model is not that it is unreal, but that it is a simpler version of the modeled thing. Likewise, IML is real, even if it could serve as a model (simpler version) of animal learning. Of course, IML is different from animal learning in myriad important respects, but it is learning nonetheless. The artificiality of IML does not make it any less learning.

A third possible criticism is that a negative answer to Q proves explanations of behavior in terms of neural processes, artificial neural-network modeling included, mistaken. However, the rationale behind this criticism is fallacious. My argument is quite straightforward: If indeed there is learning in bacteria, paramecia, plants, and inorganic machines, this only entails that *they* do not need a nervous system to learn. The possibility does not entail that learning *in animals* does not need a nervous system. Of course, the possibility of learning in spinal vertebrates indicates that some cases of learning do not require a *complete* nervous system, especially a brain, but this is not a valid reason to eschew explanations of behavior in terms of brain processes.

To be more precise, my key point here is that a negative answer to Q does not imply that nervous systems are *explanatorily irrelevant* to, *when present during, learning*. The faulty logic that drives the opposite point is the following: If some condition *A* is unnecessary for some other condition *B*, then *A* is explanatorily irrelevant to *B* when present. The obvious flaw in this logic is that the fact that *A* is unnecessary for *B* does not preclude the possibility that *A* is present with and exerts some influence on *B*. A dog barking is not necessary for sleeping: Many of us can sleep quite well *without* it. However, *if present*, it exerts a major influence on sleeping (by interrupting it, in many cases). In general, that *X* is not necessary for *Y* means that *Y* can happen without *X*, but this does not entail that *Y* will be unaffected by *X*’s presence.

The same logic applies to nervous systems, in particular, brains. Learning in spinal vertebrates indicates that some cases of vertebrate learning do not require a brain.

However, this does not mean that brains are explanatorily irrelevant *when present* during learning processes in whole, healthy vertebrates. When a brain is present, it exerts an influence on behavior that can be explained in terms of, say, a neural-network model. Of course, such a model would not apply to explanations of brainless learning in spinal vertebrates or, even more obviously, aneural organisms and inorganic machines, let alone inorganic machines. But all this only means that the model is of limited applicability, which is all right: At present, seeking a universal model that explains all forms of learning seems premature.

Acknowledgements I thank two anonymous reviewers for useful comments to a previous draft.

Compliance with Ethical Standards

Funding Information This research was not funded.

Conflict of Interest I declare no conflict of interest.

References

- Alon, U., Surette, M. G., Barkai, N., & Leibler, S. (1999). Robustness in bacterial chemotaxis. *Nature*, 397(January 14, 1999), 168–171. <https://doi.org/10.1038/16483>.
- Alpaydin, E. (2016). *Machine learning: The new AI*. Cambridge, MA: MIT Press.
- Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.
- Applewhite, P. B. (1972). Behavioral plasticity in the sensitive plant, *Mimosa*. *Behavioral Biology*, 7, 47–53. [https://doi.org/10.1016/S0091-6773\(72\)80187-1](https://doi.org/10.1016/S0091-6773(72)80187-1).
- Armus, H. L., Montgomery, A. R., & Jellison, J. L. (2006). Discrimination learning in paramecia (*P. caudatum*). *The Psychological Record*, 56, 489–498.
- Baum, W. M. (2013). What counts as behavior? The molar multiscale view. *The Behavior Analyst*, 36, 283–293.
- Bolles, R. C. (1976). Some relationships between learning and memory. In D. L. Medin, W. A. Roberts, & R. T. Davis (Eds.), *Processes of animal memory* (pp. 21–48). New York, NY: Psychology Press.
- Bunge, M., & Ardila, R. (1987). *Philosophy of psychology*. New York, NY: Springer-Verlag.
- Burgos, J. E., & Killeen, P. R. (2018). Suing for peace in the war against mentalism. *Perspective on Behavior Science*. <https://doi.org/10.1007/s40614-018-0169-2>.
- Carew, T. J., & Sahley, C. L. (1986). Invertebrate learning and memory: From behavior to molecules. *Annual Review of Neuroscience*, 9, 435–487. <https://doi.org/10.1146/annurev.ne.09.030186.002251>.
- Carew, T. J., Walters, E. T., & Kandel, E. R. (1981). Classical conditioning in a simple withdrawal reflex in *Aplysia californica*. *Journal of Neuroscience*, 1, 1426–1437. <https://doi.org/10.1523/JNEUROSCI.01-12-01426.1981>.
- Catania, A. C. (2012). *Learning* (5th ed.). Cornwall on Hudson, NY: Sloan.
- Chopin, S. F., & Buerger, A. A. (1976). Instrumental avoidance conditioning in the spinal rat. *Brain Research Bulletin*, 1, 177–183.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7–19. <https://doi.org/10.1111/1467-8284.00096>.
- Colman, A. M. (2015). *Oxford dictionary of psychology* (4th ed.). Oxford, UK: Oxford University Press.
- Cousins, L. S., Zamble, E., Tait, R. W., & Suboski, M. D. (1971). Sensory preconditioning in curarized rats. *Journal of Comparative & Physiological Psychology*, 77, 152–154. <https://doi.org/10.1037/h0031586>.
- De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, 20, 631–642. <https://doi.org/10.3758/s13423-013-0386-3>.

- Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.
- Domjan, M. (2015). *The principles of learning and behavior*. (7th ed.). Stamford, CT: Cengage Learning.
- Dowe, P. (2009). Causal process theories. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 213–233). Oxford, UK: Oxford University Press.
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Dudai, Y. (1989). *The neurobiology of memory: Concepts, findings, trends*. Oxford, UK: Oxford University Press.
- Dukas, R. (2009). Learning: Mechanisms, ecology, and evolution. In R. Dukas & J. M. Ratcliffe (Eds.), *Cognitive ecology II* (pp. 7–26). Chicago, IL: University of Chicago Press.
- Dukas, R. (2018). Cognition and learning. In A. Córdoba-Aguilar, D. González-Tokman, & I. González-Santoyo (Eds.), *Insect behavior: From mechanisms to ecological and evolutionary consequences* (pp. 267–272). Oxford, UK: Oxford University Press.
- Enquist, M., & Ghirlanda, S. (2005). *Neural networks & animal behavior*. Princeton, NJ: Princeton University Press.
- Gagliano, M., Vyazovskiy, V. V., Borbély, A. A., Grimonprez, M., & Depczynski, M. (2016). Learning by association in plants. *Nature Scientific Reports*, 6, 38427. <https://doi.org/10.1038/srep38427>.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gallistel, C. R., & King, A. P. (2010). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Malden, MA: Wiley-Blackwell.
- Gori, M. (2018). *Machine learning: A constraint-based approach*. Cambridge, MA: Morgan Kauffman.
- Grau, J. W. (2014). Learning from the spinal cord: How the study of spinal cord plasticity informs our view of learning. *Neurobiology of Learning & Memory*, 108, 155–171. <https://doi.org/10.1016/j.nlm.2013.08.003>.
- Grau, J. W., & Joynes, R. L. (2005). Neurofunctionalism revisited: Learning is more than you think it is. *International Journal of Comparative Psychology*, 18, 46–59 Retrieved from <https://escholarship.org/uc/item/2ps1d9q5>.
- Haselgrove, M. (2016). *Learning: A very short introduction*. Oxford, UK: Oxford University Press.
- Hennessey, T. M., Rucker, W. B., & McDiarmid, C. G. (1979). Classical conditioning in paramecia. *Animal Learning & Behavior*, 7, 419–423. <https://doi.org/10.3758/BF03209695>.
- Higgins, E. S., & George, M. S. (2013). *The neuroscience of clinical psychiatry: The pathophysiology of behavior and mental illness* (2nd ed.). Philadelphia, PA: Lippincott, Williams, & Wilkins.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan.
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York, NY: Appleton-Century-Crofts.
- Jakob, E., Skow, C., & Long, S. (2011). Plasticity, learning and cognition. In M. E. Herberstein (Ed.), *Spider behaviour: Flexibility and versatility* (pp. 307–347). Cambridge, UK: Cambridge University Press.
- Kearney, A. J. (2015). *Understanding applied behavior analysis: An introduction to ABA for parents, teachers, and other scholars* (2nd ed.). London, UK: Kingsley.
- Kimble, G. A. (1961). *Hilgard and Marquis' conditioning and learning* (2nd ed.). New York, NY: Appleton-Century-Crofts.
- Lachman, S. J. (1997). Learning is a process: Toward an improved definition of learning. *Journal of Psychology*, 131, 477–480. <https://doi.org/10.1080/00223989709603535>.
- Langley, P. (1996). *Elements of machine learning*. San Francisco, CA: Morgan Kauffman.
- Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. Berkeley & Los Angeles: University of California Press.
- Levitis, D. A., Lidicker Jr., W. Z., & Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal Behaviour*, 78, 103–110. <https://dx.doi.org/10.1016/j.anbehav.2009.03.018>.
- Machado, A. (2005). Experimental methods and conceptual confusion. *International Journal of Comparative Psychology*, 18, 28–33 Retrieved from <https://escholarship.org/uc/item/40v191xn>.
- Mangel, M. (1993). Motivation, learning, and motivated learning. In D. R. Papaj & A. C. Lewis (Eds.), *Insect learning: Ecological and evolutionary perspectives* (pp. 158–173). New York, NY: Routledge.
- Marsland, S. (2009). Using habituation in machine learning. *Neurobiology of Learning & Memory*, 92, 260–266.
- Matthews, R. W., & Matthews, J. R. (2010). *Insect behavior* (2nd ed.). London, UK: Springer.
- Mazur, J. E. (2017). *Learning and behavior* (8th ed.). New York, NY: Routledge.
- Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin*, 85, 1256–1274. <https://doi.org/10.1037/0033-2909.85.6.1256>.
- McConnell, J. V. (1966). Comparative physiology: Learning in invertebrates. *Annual Review of Physiology*, 28, 107–136. <https://doi.org/10.1146/annurev.ph.28.030166.000543>.

- Menzel, R., Greggers, U., & Hammer, M. (1993). Functional organization of appetitive learning and memory in a generalist pollinator, the honey bee. In D. R. Papaj & A. C. Lewis (Eds.), *Insect learning: Ecological and evolutionary perspectives* (pp. 79–125). Dordrecht, The Netherlands: Springer.
- Millikan, R. G. (1993). What is behavior? A philosophical essay on ethology and individualism in psychology, Part 1. In R. Millikan (Ed.), *White queen psychology and other essays for Alice* (pp. 135–150). Cambridge, MA: MIT Press.
- Mitchell, T. M. (1997). *Machine learning*. New York, NY: McGraw-Hill.
- Olson, M. H., & Hergenhahn, B. R. (2016). *An introduction to theories of learning* (9th ed.). New York, NY: Routledge.
- Papaj, D. R., & Lewis, A. C. (Eds.). (1993). *Insect learning: Ecology and evolutionary perspectives*. New York, NY: Routledge.
- Pear, J. J. (2016). *The science of learning* (2nd ed.). New York, NY: Routledge.
- Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics*, 37, 1–38.
- Pierce, W. D., & Cheney, C. D. (2017). *Behavior analysis and learning: A biobehavioral approach* (6th ed.). New York, NY: Routledge.
- Reilly, S., & Schachtman, T. R. (2005). Pavlovian conditioning requires ruling out nonassociative factors to claim conditioning occurred. *International Journal of Comparative Psychology*, 18, 34–37 Retrieved from <https://escholarship.org/uc/item/6455042j>.
- Rescher, N. (1996). *Process metaphysics: An introduction to process philosophy*. Albany, NY: State University of New York Press.
- Rescorla, R. A. (1988). Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience*, 11, 329–352. <https://doi.org/10.1146/annurev.ne.11.030188.001553>.
- Sahley, C. L., Rudy, J. W., & Gelperin, A. (1984). Associative learning in a mollusk: a comparative analysis. In D. L. Alkon & J. Farley (Eds.), *Primary neural substrates of learning and behavior change* (pp. 243–258). Cambridge, UK: Cambridge University Press.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- Salmon, W. (1998). *Causality and explanation*. New York, NY: Oxford University Press.
- Shurrager, P. S., & Culler, E. (1940). Conditioning in the spinal dog. *Journal of Experimental Psychology*, 26, 133–159. Retrieved from <https://doi.org/10.1037/h0054950>
- Simon, H. A. (1983). Why should machines learn? In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 25–37). New York, NY: Springer.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, 57, 193–216. <https://doi.org/10.1037/h0054367>.
- Teuscher, C. (2002). *Turing's connectionism: An investigation of neural network architectures*. London, UK: Springer-Verlag.
- Thorndike, E. L. (1913). *The psychology of learning*. New York, NY: Columbia University.
- Thorpe, W. H. (1943). Types of learning in insects and other arthropods. *British Journal of Psychology*, 33, 220–234. <https://doi.org/10.1111/j.2044-8295.1943.tb01059.x>.
- Wichterman, R. (1986). *The biology of paramecium*. New York, NY: Plenum.
- Yi, T.-M., Huang, Y., Simon, M. I., & Doyle, J. (2000). Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Science of the United States of America*, 97, 4649–4653. <https://doi.org/10.1073/pnas.97.9.4649>.