Learning in Individual Organisms, Genes, Machines, and Groups: A New Way of Defining and Relating Learning in Different Systems

Jan De Houwer & Sean Hughes

Ghent University

In Press. Perspectives on Psychological Science.

Authors Note

JDH and SH, Department of Experimental Clinical and Health Psychology, Ghent University. The preparation of this paper was supported by Grant BOF16/MET_V/002 of Ghent University to JDH. This paper contains parts from a previous, unpublished paper by Hughes and De Houwer (2020). Thanks to Adelheid Soubry for discussions regarding genetic learning and to Yannick Boddez, Baptist Liefooghe, and two anonymous reviewers for comments on an earlier draft of this paper. Correspondence concerning this paper can be sent to jan.dehouwer@ugent.be .

Abstract

Learning is a central concept in many scientific disciplines. Communication about research on learning is, however, hampered by the fact that different researchers define learning in different ways. In this paper, we introduce the extended functional definition of learning that can be used across scientific disciplines. We provide examples of how the definition can be applied to individual organisms, genes, machines, and groups. Using the extended functional definition (a) reveals a heuristic framework for research that can be applied across scientific disciplines, (b) allows researchers to engage in intersystem analyses that relate the behavior and learning of different systems, and (c) clarifies how learning differs from other phenomena such as (changes in) behavior, damaging systems, and programming systems.

Keywords: learning, behavior, conceptual analysis, levels of analysis

Learning in Individual Organisms, Genes, Machines, and Groups:

A New Way of Defining and Relating Learning in Different Systems

Learning is of fundamental interest to many scientists. Psychologists have devoted decades of effort to understand how humans learn while their counterparts in sociology, economics, and anthropology examine how groups of humans and even societies learn. Within the biological sciences, researchers analyze learning throughout the animal kingdom, from whole organisms down to their neurons, and even their genes, while computer scientists probe learning in robots and other artificial systems. Despite this common interest in learning, there is little agreement on what it means to say that learning has taken place. For instance, Barron et al. (2015) list no less than 59 different definitions of learning that originate from different scientific disciplines such as psychology, neuroscience, and computer science (also see Burgos, 2018; Lachman, 1997). As they correctly point out, this divergence of definitions impairs transdisciplinary collaboration and synthesis. Although it is naïve to think that there is one “true” definition of learning, we believe that it is worthwhile to strive for definitions of learning that can be used across different disciplines. Such definitions would foster communication between researchers and thereby facilitate cross-fertilization between disciplines. In this paper, we propose a definition of learning that can be used to describe learning in different systems such as whole individual organisms, parts of individuals, non-organic machines, as well as groups.

The definition that we put forward is an extension of the functional definition of learning provided by De Houwer et al. (2013). Hence, we refer to our definition as the extended functional definition of learning. In Part I of this paper, we first explain how it goes beyond the definition of De Houwer et al (2013) and then specify the concepts that it builds on. To make the definition more concrete, we apply it to several known types of learning in whole individual organisms (e.g., a human, dog, or rat). In Part II, we illustrate how the extended functional definition of learning can also be used to describe learning in other systems, more specifically genes, machines, and groups. In Part III, we argue that the extended functional definition of learning reveals a common heuristic framework for research on learning that can be applied to different systems. As such, it stimulates cross-disciplinary interactions and highlights issues to be addressed in future research. It also allows for intersystem analyses about how behavior and learning in one system is conditional on behavior or learning in another system. Finally, it allows for more conceptual precision by clarifying the differences between closely related concepts such as behavior, learning, and programming. In Part IV, we discuss possible limitations of the definition, including the fact that using it requires assumptions about causality that can be difficult to verify.

Throughout the paper, we use concrete examples to illustrate abstract ideas. Some of these examples correspond to known phenomena. Others are hypothetical. Please keep in mind that we do not wish to make claims about the validity of the specific examples that we selected, nor do we wish to limit our ideas to those examples. Instead, the examples are simply meant illustrate the nature and merits of our definition of learning.

Part I: Defining Learning

The Extended Functional Definition of Learning

De Houwer et al. (2013) defined learning as "*changes in the behavior of an organism that are the result of regularities in the environment of that organism"* (p. 633) and noted that "*the concept of regularity encompasses all states in the environment of the organism that entail more than the presence of a single stimulus or behavior at a single moment in time*" (p. 634). This definition qualifies as a functional definition in that changes in behavior are said to be a function of regularities in the environment. Although it has proven useful within psychology (see De Houwer & Hughes, 2020, for an overview), Burgos (2018) correctly pointed out that the functional definition of De Houwer et al. explicitly refers to individual organisms and therefore does not apply to others systems such as inorganic machines. Another limitation is that De Houwer et al. (2013), like many others who proposed a definition of learning, did not provide a precise definition of the concept “behavior”. They only noted that behavior “encompasses every observable response that a living organism can make” (p. 633) without clarifying what they mean with a response.

To remedy these limitations, we extend the definition of De Houwer et al. (2013) by defining learning as “*changes in the way a system behaves towards a stimulus that are the result of regularities in the environment of that system*”. A system can refer not only to a whole organism (e.g., a human, dog, or rat) but also to inorganic, man-made machines (e.g., robots, self-driving cars, artificial intelligence), to constituent parts of such systems (e.g., the organs, neurons, or genes of an individual organism), or to groups (e.g., herds, colonies, societies). More generally, a system can be anything non-static, that is, anything that can be in different states. A state can be defined as a condition or way of being of a system. This is a broad definition which includes both static states (e.g., being at a certain location) and dynamic states (e.g., moving in a certain direction) and which applies to both objective states (i.e., that can be observed by researchers who analyze a system; e.g., moving in a certain direction) and subjective states (i.e., that can be observed only by the system itself; e.g., conscious feelings and thoughts). A state transition occurs when the state of a system changes. We define behavior as “*transitions in the state of a system that are due to a stimulus in the environment*”. Hence, also our definition of behavior is a functional definition in that the state of a system is said to be a function of a stimulus in the environment. We thus think of behavior as a response of the system to a stimulus in the environment and will use the expression “responding to” as a synonym for “behaving”. This implies that changes in behavior towards a stimulus are changes in the way a particular stimulus influences the state of a system and that learning is the impact of regularities on how a particular stimulus changes the state of a system.

How Do Our Definitions Relate to Other Definitions?

To avoid misunderstandings, it is important to point out that our definitions of regularities, states, behavior, and learning deviate from how these concepts are sometimes used by other scientists or lay persons. First, whereas many might think of regularities as necessarily involving events that occur repeatedly, we adopt the broader definition of regularities that encompasses all events that are orderly, even if those events occur at one point in time. Hence, also the co-occurrence of two or more stimuli at one point in time qualifies as a regularity from this perspective (also see De Houwer et al., 2013; De Houwer & Hughes, 2020, Introduction, Footnote 2).

Second, some might think of states as necessarily static. As noted above, we also consider dynamic ways of being as states of a system (e.g., being in a state of moving forward). We realize that this can create confusion between the notion of dynamic states and the notion of state transitions because a dynamic state can itself be described as consisting of transitions in states (e.g., a movement as a transition from being at one location to being at another location). Nevertheless, by allowing for dynamic states, we broaden the class of possible state transitions so that it includes also transitions from a static state to a dynamic state, from a dynamic state to a static state, and from one dynamic state to another dynamic state. [1]

[1] It is worth noting that there is merit in thinking about state transitions in a probabilistic manner (e.g., there is a higher or lower probability of the system being observed to transition from one state to another at a given moment in time). For ease of communication, however, we will talk about state transitions that do or do not occur at a given moment in time.

This brings us to our third concept: behavior. Often the term "behavior" is used to refer to what we would call a dynamic state of a system: a way of being that involves state transitions (e.g., moving forward). In contrast, we think of "behaving" as "responding to". We thus use the term behavior in a more restricted manner that requires both a state transition and an assumption about the functional cause of the state transition. For instance, whereas the claim that a system is in the state of moving forward (i.e., transitioning from being in location A to being in location B) implies only a description of the system, referring to moving forward as a behavior requires the assumption that there is a particular stimulus that caused the state transition, that is, a stimulus without which this state transition would not have occurred (e.g., the stimulus that the system approaches or the stimulus that the system moves away from). Hence, state transitions that occur spontaneously (i.e., that are not a function of stimuli in the current or past environment of the system) do not qualify as behavior in the way we define it.

Finally, unlike many other definitions of learning (e.g., learning as the storage of information in the brain; see Barron et al., 2015, for a review), our definition does not imply assumptions about representations or mechanisms. This does not mean that using the definition requires a denial of the existence of representations or mechanisms via which information is processed. It only means that the definition allows one to identify instances of learning and describe the conditions under which learning occurs without making reference to the existence or nature of representations and mechanisms that mediate learning. We believe that this approach is particularly advantageous when studying different systems that probably differ substantially with regard to the representations and mechanisms that mediate learning. The extended functional definition of learning only requires assumptions about the causal impact of regularities on changes in behavior towards a specific stimulus. We use the term "causal" in the sense of "functional causation" which implies only that an outcome is a function of (i.e., would not have been that way without) the cause. Although functional causes do not inform us about the representations and mechanism via which a cause produces an outcome, the claim that an outcome is a function of a cause does provide information about the determinants of the outcome (see Chiesa, 1992). Functional concepts therefore provide more than mere description. By specifying what causes what, functional definitions also set the stage for research on the representations and mechanisms that mediate a causal relation (De Houwer & Hughes, 2020).

Please note that neither our functional concepts (i.e., behavior, learning), nor our descriptive concepts (i.e., regularities, systems, stimuli, states, state transitions) require ontological assumptions. We simply believe that, as a researcher, it can be useful to formulate assumptions about (amongst other things) systems that are present in the world, the states that those systems can be in, the presence of stimuli and regularities in the environment of those systems, and the functions of those stimuli and regularities (e.g., their function to alter the state of a system or to influence the way in which a stimulus alters the state of a system). In order to conduct research, it is not necessary that those assumptions are correct in an ontological sense. It suffices that they help the researcher to reach his or her scientific goals (e.g., to predict and influence behavior; see Hayes & Brownstein, 1985). This also implies that the scientific goals of the researcher determine which system, states of the system, stimuli, regularities, and functions that the researcher focusses on.

Applying the Extended Functional Definition of Learning to Known Learning Phenomena in Whole Individual Organisms

In research on learning in individual whole organisms, several types of learning can be distinguished based on the type of regularity that produces the change in behavior toward a stimulus (see De Houwer & Hughes, 2020, for an overview). In this section, we consider three of these types of learning. Arguably, the most basic type of learning involves a change in behavior that is due to the *repeated presentation of one stimulus* (see Figure 1). To illustrate, imagine that you (i.e., the system) suddenly hear a loud bang. At this time (Time 1), the loud bang (stimulus) may cause you to transition from being in calm state to being in a highly startled state. This state transition as the result of the loud bang constitutes an instance of behavior (i.e., Behavior 1 at Time 1). Now imagine that you repeatedly experience the loud bang. This constitutes a regularity in that it involves more than one stimulus at one point in time. When you afterwards (at Time 2) experience the loud bang again, you may be startled to a far lesser extent than before. This means that the loud bang caused a different state transition at Time 1 than at Time 2 and thus that a change in behavior towards the loud bang has occurred (i.e., responding to the loud bang is different at Time 1 than at Time 2). If it can be shown that this change in behavior is due to the regularity in the presence of the loud bang (e.g., by showing that the change does not occur without the regularity) then it can be concluded that learning occurred. Common examples of this type of learning are *habituation* (what we have just described; see Figure 1) and *sensitization* (an increase in the intensity of a response to a stimulus due to the repeated presentation of the stimulus; see De Houwer & Hughes, 2020, Chapter 1).

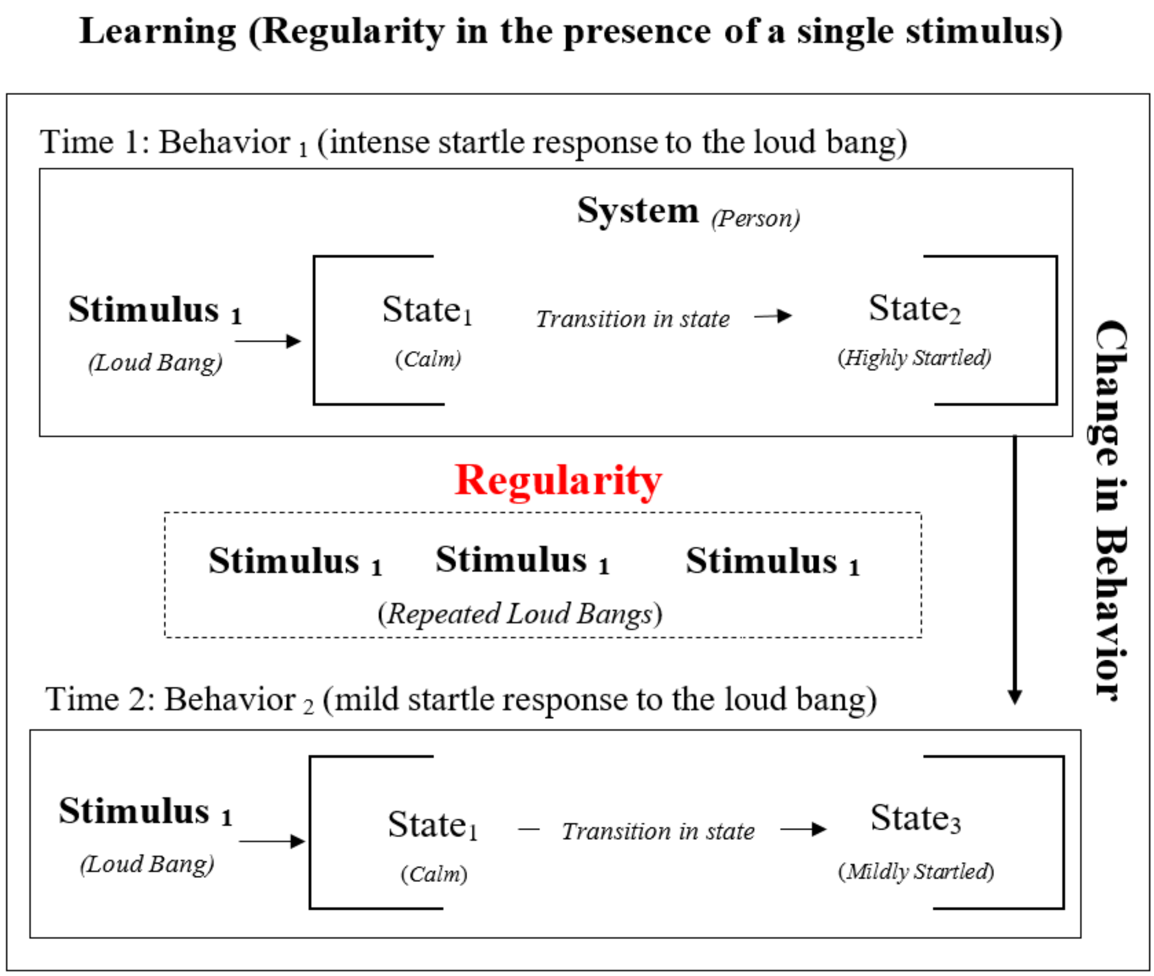

Figure 1. A visual illustration of learning that involves a regularity in the presence of a single stimulus (habituation).

A second type of learning involves a change in behavior that is due to a *regularity in the presence of two stimuli*. This type of learning can be referred to as classical or Pavlovian conditioning (see Bouton, 2016, for an overview). To illustrate, imagine that we ring a bell (Stimulus 1) in the presence of a dog (system). At Time 1, the bell does not lead to a particular state transition (e.g., no salivation response to the bell). Later on we introduce a regularity involving two stimuli: we pair the ringing of the bell with the administration of food. At Time 2, the bell (Stimulus $_1$) now leads to the state transition we are interested in (i.e., salivation response to the bell). This increase in the salivation response to the bell from Time 1 to 2 constitutes a change in behavior. If we can show that this change is due to the regularity between the ringing of the bell and food (and not, for instance, to the mere presence of food), then we can say that learning, more specifically classical conditioning, has taken place (see Figure 2).

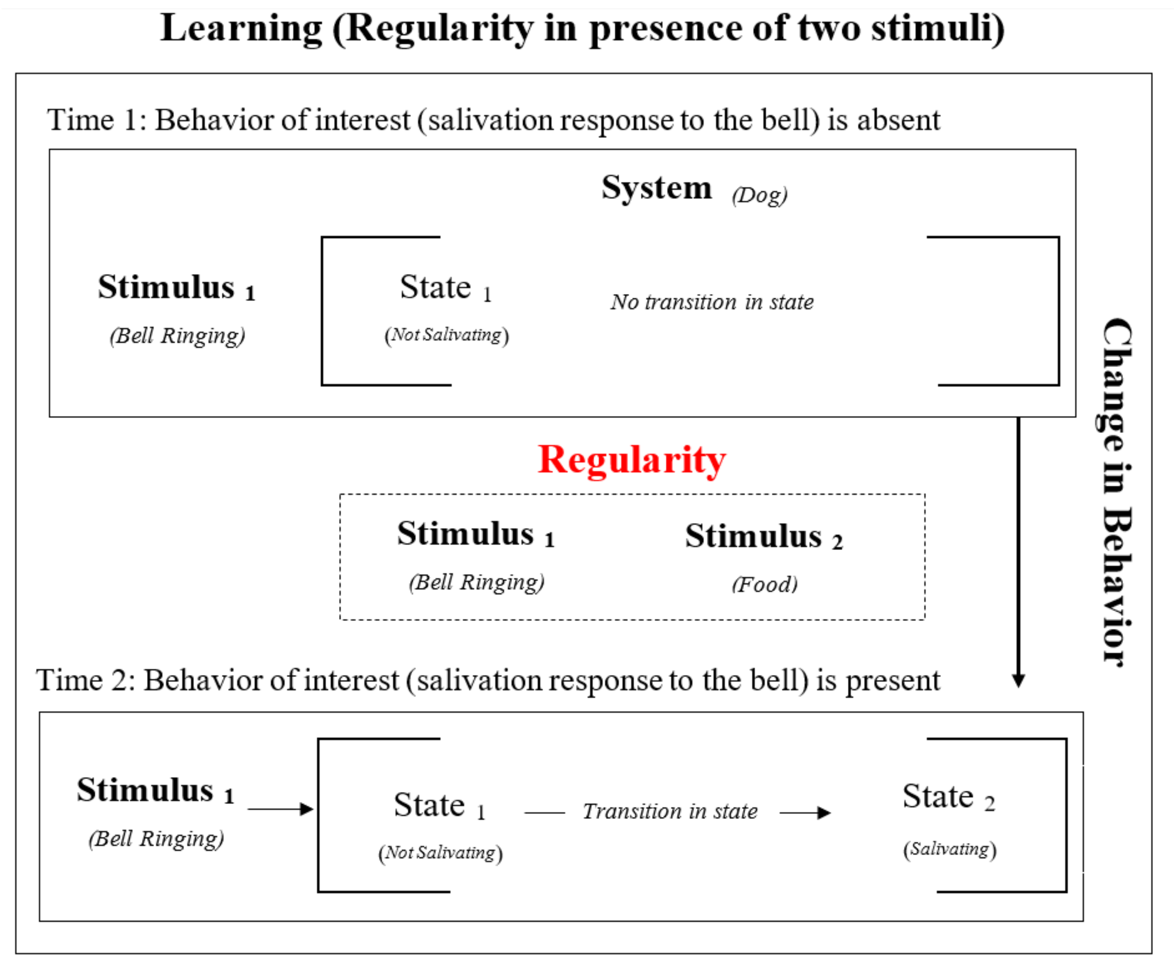

Figure 2. A visual illustration of learning that involves a regularity in the presence of two stimuli (classical conditioning).

A third type of learning, which can be referred to as operant conditioning, involves a change in behavior that is due to a *regularity in the presence of stimuli and responses* (see Catania, 2013, for a review). To illustrate, imagine that we place a rat (system) into a training chamber which contains a light, a lever which can be pressed, and a food dispenser. At Time 1, turning on the light in the chamber does not lead to the behavior we are interested in (i.e., lever pressing in response to the light). We then introduce a regularity such that if, in the presence of the light (Stimulus 1), the behavior does occur (e.g., because the rat is actively prompted by the researcher into pressing the lever) then food is delivered (Stimulus 2). At

Time 2, we now observe that the presence of the light is followed by lever pressing. Hence, we can infer that the behavior of lever pressing in response to the light has changed from Time 1 to Time 2. If this change is due to a regularity involving stimuli and responses (i.e., the fact that lever pressing leads to food when the light is on), then the change can be referred to as learning, more specifically as an operant conditioning effect (see Figure 3). [2]

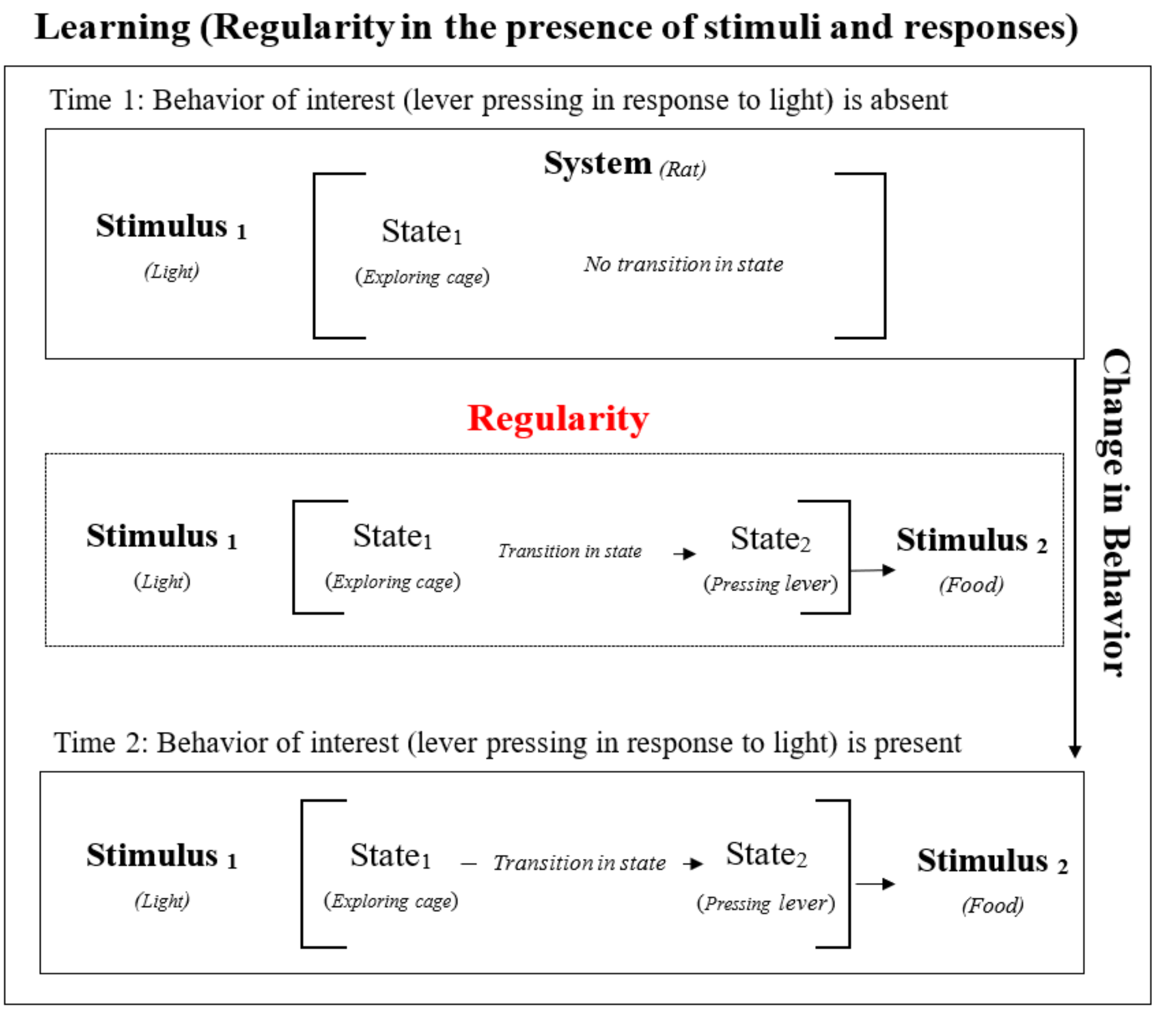

Figure 3. A visual illustration of learning that involves a regularity in the presence of stimuli and responses (operant conditioning).

[2] Note that cage exploration and lever pressing are dynamic states that in all likelihood also qualify as instances of behavior: the state transitions involved in these dynamic states are almost certainly not random but a function of stimuli in the environment of the rat (e.g., lever pressing is a function of the presence of a lever). For our analysis, however, it does not matter whether cage exploration and lever pressing are behaviors or mere dynamic states. Because we are interested only in behavior towards the light, we need to focus only on the impact of the light on the state of the system. Because the impact of the light on the system is different at Time 1 (i.e., no impact) than at Time 2 (i.e., light causes a transition from cage exploration to lever pressing), we can say that there is a change in the behavior towards the light. If this change in behavior towards the light is due to a regularity, then this change qualifies as an instance of learning.

Part II: Learning Beyond the Individual Organism

In Part I we focused our attention on learning in one type of system – individual whole living organisms such as humans, rats, and dogs. However, the extended functional definition of learning also applies to many other systems. Applying the definition to these other systems simply involves specifying the nature of the system, possible states and behaviors of the system (i.e., how the state of the system transitions as the result of stimuli in the environment), possible changes in the behavior of the system, and regularities that might cause those changes in behavior. In what follows, we briefly illustrate how the definition applies to three systems other than whole individual organisms: genes, machines, and groups of individual organisms.

Genetic Learning

All living organisms contain genetic material. Although there are debates about the exact nature of genes, one could argue that genes are subsets of this genetic material, next to other parts of DNA and other genetic material in the nucleus, mitochondria, and cytoplasm (Gerstein et al., 2007). Much of this other genetic material seems to be involved in the epigenetic regulation of genes (e.g., via methylation; Moore, 2015). Both genes and regulatory genetic material can be in different states. For instance, genes can be inactive or actively involved in the production of proteins. Likewise, epigenetic regulatory mechanisms can be inactive or actively involved in the regulation of genes. Hence, at least in principle, genes and epigenetic regulatory mechanisms can behave (i.e., transition their state in response to a stimulus), change their behavior (i.e., respond differently to the same stimulus at different points in time), and learn (i.e., change their behavior as the result of regularities).

To illustrate, consider the idea of “stress memory” in plants. When certain varieties of maize plants are subjected to the same stimulus (dehydration) over and over again they show improved retention of water compared to the first time they encountered that stimulus (e.g., Ding et al., 2012; Virlouvet & Fromm, 2014). These changes in behavior in the face of repeated contact with the stimulus are evident not only at the organismic and cellular levels, but at the genetic level as well: transcription levels of stress response genes are significantly higher after multiple presentations of a stressful stimulus (dehydration) compared to the first presentation of that stimulus (for more on “transcriptional stress memory”, see Avramova, 2015).

From our perspective, the state transition (from no transcription to a low level of transcription of stress response genes) produced by the stimulus (dehydration) at Time 1 is an example of a genetic behavior. So too is the state transition (from no transcription to a high level of transcription of stress response genes) that occurs when the same stimulus (dehydration) is presented at Time 2. This *increase* in transcription levels from Time 1 to 2 constitutes a change in genetic behavior, that is, a change in how the stress response genes respond to dehydration. If this change in behavior is due to the repeated occurrence of dehydration, genetic learning can be said to have taken place. In this case, it would qualify as an instance of *genetic sensitization* (see Figure 5). If the intensity of gene expression in response to dehydration were to decrease as a function of repeated dehydration then we would instead speak of *genetic habituation* (also see Hughes & De Houwer, 2020). [3]

[3] The role of royal jelly in the development of queen bees also seems like a candidate for this type of learning (see Chittka & Chittka, 2010; Kucharski, 2008). A change in behavior towards the presence of royal jelly (decrease in DNMT3 expression and the associated upregulation of developmental genes in queen-destined larvae) appears to take place when the system is exposed to the royal jelly over and over again. Note that testing this claim would require that we systematically manipulate the number of times a system (honey bee larvae) is exposed to the same stimulus (royal jelly) while controlling for possible confounding factors (e.g., mere passage of time) and that we look for differences in genetic responding to the administration of royal jelly.

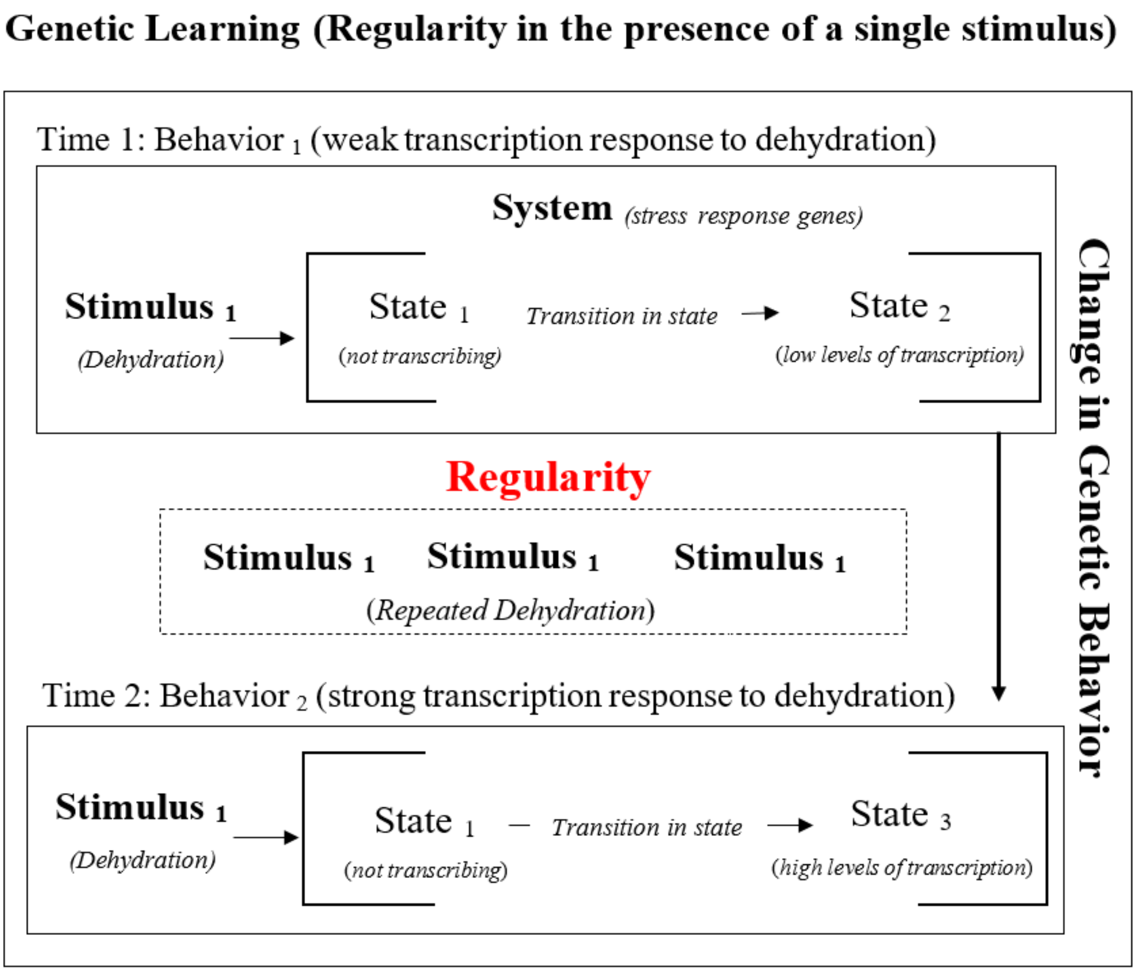

Figure 4. A visual illustration of one instance of genetic learning (i.e., genetic sensitization).

Learning in Inorganic Machines

Although one could argue that living organisms are machines, a distinction is often made between living organisms and inorganic machines (in a broad sense that includes neural nets and artificial intelligence systems). The idea that also inorganic machines can learn has given rise to an extensive literature on machine learning (e.g., Burgos, 2018; Rahwan et al., 2019). Here we consider machine learning from the perspective of the extended functional definition of learning.

Like individual organisms, machines can be in different states. For instance, a robot can be static, it can move forward, or it can change direction in a particular way (e.g., turn left, turn right). Likewise, an artificial intelligence (e.g., a neural net) can signal to move a chess piece to the right or to the left. Hence, also for machines it makes sense to ask whether they can behave (i.e., transition their state as the result of a stimulus in the environment), change their behavior towards specific stimuli (i.e., change the way they transition their state as the result of a particular stimulus in the environment), and learn (i.e., change their behavior towards a specific stimulus as the result of a regularity in their environment).

Consider the example of a humanoid robot that is being trained to navigate the world and circumvent obstacles. From the perspective of the extended functional definition of learning, the claim that a robot learns to circumvent an obstacle would imply the following (see Figure 5). First, at Time 1, when the robot approaches the location of the obstacle (Stimulus 1), it simply does what it would have done without the obstacle: it moves forward and thus fails to go beyond the obstacle. Now imagine that during subsequent attempts, the robot is prompted into moving to the left of the obstacle once nears the obstacle. Because the robot turns left, it succeeds in circumventing the obstacle. At Time 2, we see that the robot successfully moves via the left of the obstacle even without being prompted. This means that the robot has changed its behavior towards the obstacle: it did not change direction when nearing the obstacle at Time 1 but it did change direction when nearing the same obstacle at Time 2. If this change in behavior is due to the regularities that occur in between Time 1 and Time 2 (i.e., because it succeeds in moving beyond the obstacle when it turns left after detection of the obstacle), then we can say that the robot has learned, more specifically, it has learned to circumvent the obstacle. Because the regularities that cause the change in behavior involve stimuli and responses, this change in behavior qualifies as an instance of operant conditioning of a humanoid robot.

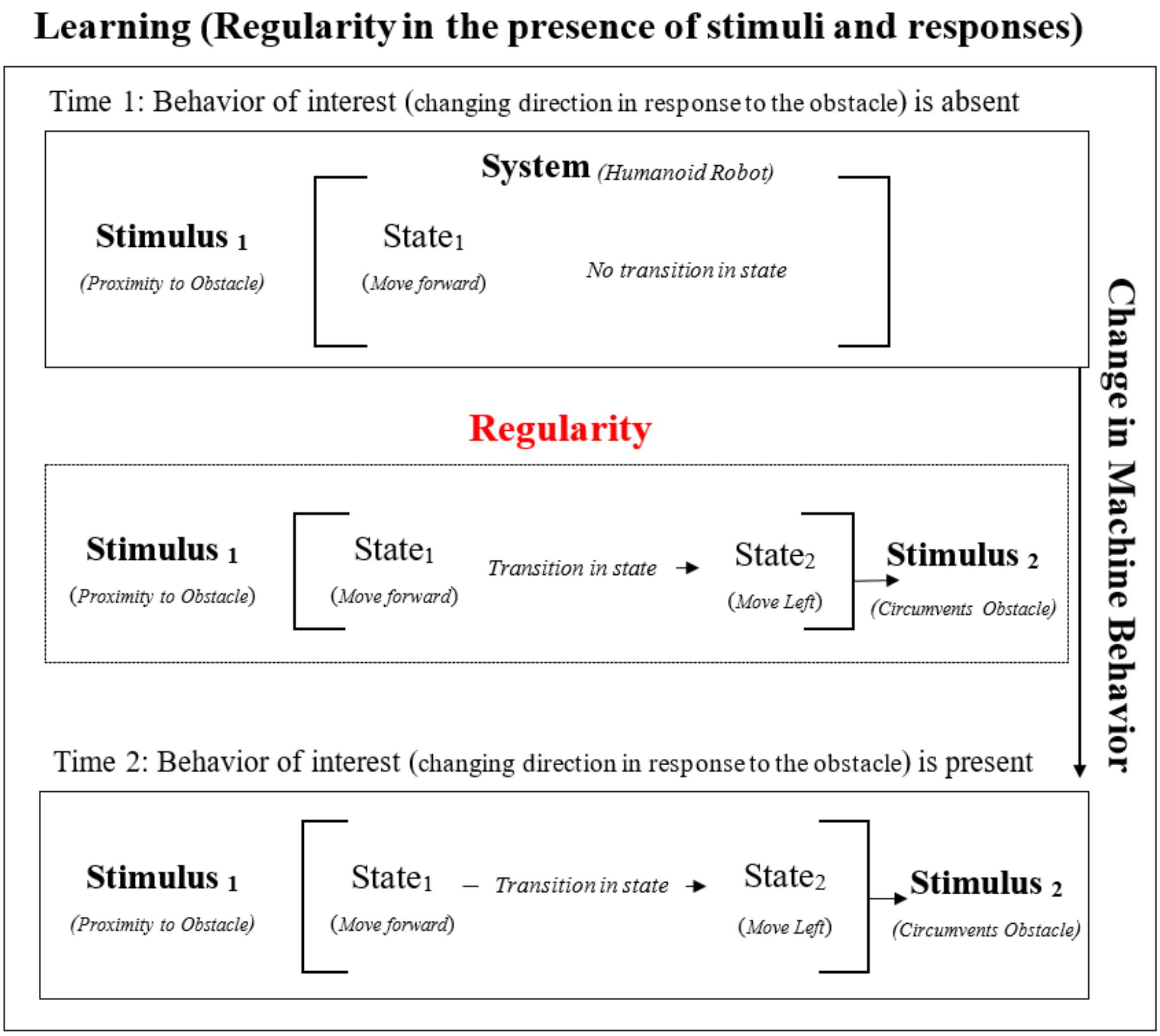

Figure 5. A visual illustration of one type of machine learning (i.e., operant conditioning in a humanoid robot).

Group Learning

Although the term “group” may seem intuitive in everyday language, there is much debate about what the term means in science (Pietraszewski, 2021). Regardless of this debate, it seems safe to say that groups always involve the presence of multiple group members (e.g., organisms, machines). From the perspective of the extended functional definition of learning, each group member could be treated as a separate system but all group members together can also be treated as a single system. Whereas the former would involve analyzing the states and behaviors of a group member, the latter would focus on the states and behaviors of the group as a whole. For instance, a school of fish can be in a state of moving forward, turning left, or turning right. Hence, it makes sense to ask whether a group can behave (transition its state in response to a stimulus), change its behavior towards a specific stimulus (show different state transitions in response to the same stimulus at different moments in time), and learn (change its behavior towards a specific stimulus as the result of a regularity in its environment).

Consider the hypothetical example of a school of fish that is often found in the vicinity of a shipwreck. At Time 1, you observe that when the school is passing the shipwreck at a certain distance, it changes direction and approaches the shipwreck. At a later point in time, however, a predator (e.g., a shark) starts living near the shipwreck. From then onwards, whenever the school of fish changes direction to approach the shipwreck, it is attacked by the predator. At a later Time 2, you notice that when the school is passing the shipwreck, it no longer changes direction to approach the shipwreck. The school thus changed its behavior towards the shipwreck: it responds differently to the presence of the shipwreck at Time 1 than at Time 2. If this change of behavior occurred because of the regularity that is present in the environment between Time 1 and Time 2 (a predator attack when the school approaches the shipwreck), then this change qualifies as in instance of learning. Because this regularity involves stimuli and behaviors, the change in the behavior of the school qualifies as an instance of operant conditioning.

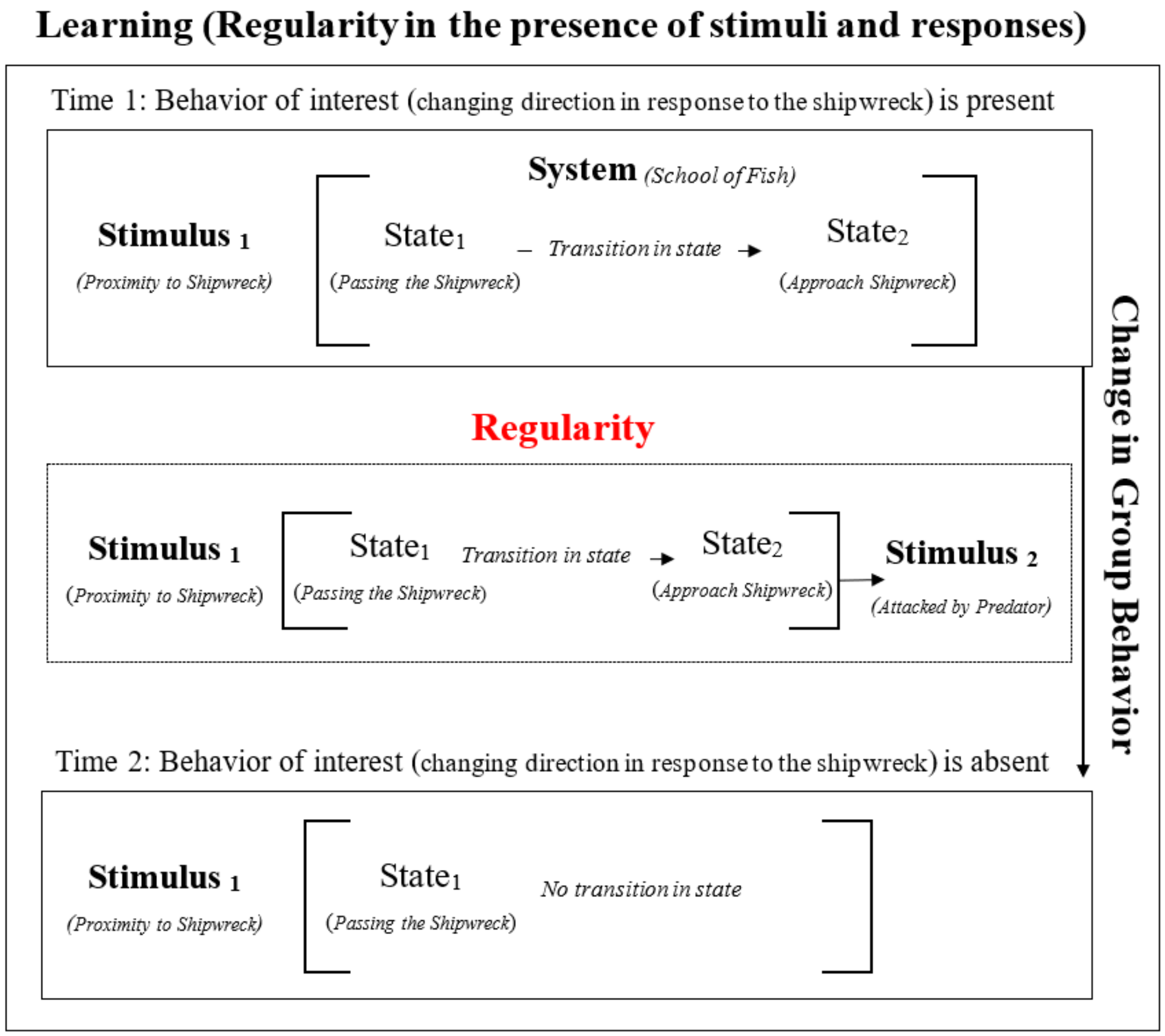

Figure 6. A visual illustration of one type of group learning (i.e., operant conditioning of a school of fish).

Part III: Heuristic and Generative Value of our Definition

So far we have seen how our extended functional definition can be applied across many types of systems. But what exactly is the value of having such a widely applicable definition of learning? In this section of the paper, we first illustrate how the definition can facilitate the study of learning in one type of system (e.g., genes) by relating it to research on learning in another type of system (e.g., whole individual organisms). We then discuss how the definition allows for intersystem analyses about how behavior and learning in one system (e.g., the algorithm in a humanoid robot) is conditional on learning in another system (e.g., the robot). Finally, we clarify how the definition allows us to tease apart closely related concepts such as behavior, learning, and programming, thereby allowing for more precise communication between researchers.

Increasing Knowledge within a Scientific Discipline by Looking at Other Scientific Disciplines: A Cross-disciplinary Heuristic Framework for Research on Learning

As Barron et al. (2015) correctly pointed out, the fact that researchers from different disciplines use different definitions of learning, impedes transdisciplinary collaboration. Even though there might be some common ground in the various definitions (see Barron et al., 2015, for suggestions), this does not change the fact that different definitions often imply different criteria for establishing whether (a specific type of) learning has occurred (see Hughes & De Houwer, 2020, for examples). Because the extended functional definition can be adopted in various scientific disciplines, it could be used to improve transdisciplinary collaboration in research on learning. Adopting the extended functional definition of learning across disciplines would not only allow for consistency in communicating about learning, but also puts forward shared criteria for establishing (types of) learning. This in turn provides the basis for an exchange of ideas between disciplines where research in one discipline can help organize existing research and provide inspiration for future research in another discipline.

The examples provided in Part II already illustrate how concepts such as habituation and operant conditioning can be used in different disciplines that study different types of systems. Via this conceptual bridge, one can verify whether variables that are known to moderate habituation and operant conditioning in one type of system have similar effects on habituation and operant conditioning in another type of system. For instance, we know that habituation in whole individual organisms depends on whether stimulus presentations are massed or distributed over time (see Thompson, 2009, for a review). To the best of our knowledge, it has not yet been examined whether this variable has a similar impact on habituation in genes. Examining this issue can provide new methods for predicting and influencing the behavior and learning of genes and epigenetic mechanisms. This simple example illustrates how the shared use of the extended functional definition of learning across disciplines can help to quickly identify gaps in our knowledge.

When the heuristic framework is used across disciplines, it not only helps highlight gaps in our knowledge but also reveals similarities and differences between the moderators of learning in different systems. By promoting knowledge of the moderators of learning, the framework helps researchers in various disciplines to influence learning in the system they focus on (i.e., to influence whether and to which extent it occurs) and to test theories about the underlying processes that allow for learning occur (i.e., to examine whether theories correctly predict when learning will occur; see De Houwer & Hughes, 2020).

Increasing Knowledge about Interactions between Systems by Relating Behavior and Learning in Different Systems: Intersystem Analyses

Although it is possible to distinguish between different types of systems, often those systems are interrelated. For instance, an individual fish can be part of a school of fish. Likewise, an artificial intelligence is part of a humanoid robot. Or genetic material and neurons can be part of a whole individual organism. In these cases, an analysis of the behavior or learning in one system can be related to an analysis of the behavior or learning in another system. Because the extended functional definition allows one to analyze behavior and learning in different systems using the same concepts, it facilitates the analysis of interactions between behavior and learning in different systems. In this section, we provide three examples of such intersystem analyses.

First, consider a humanoid robot than contains an artificial intelligence, more specifically, an algorithm that changes weights in an artificial neural net that controls the behavior of the robot. Imagine that when the robot manages to circumvent the obstacle by turning left, the algorithm strengthens the links between nodes that code the sensory input observed at that location with nodes that code the motor responses for turning left. Each time the robot manages to overcome the obstacle, these links are strengthened to the same extent. Eventually, after a certain number of times that the robot has circumvented the obstacle, the link in the network has become so strong that the robot will immediately change its direction (i.e., move to the left of the obstacle) upon detection of the obstacle.

From the perspective of the extended functional definition of learning, one can say that, in the above example, the learning of the humanoid robot is conditional on the behavior of the algorithm. The algorithm merely *behaves*: it changes its state from inactive (i.e., not strengthening links) to active (i.e., strengthening links) as the result of an outcome (i.e., circumventing an obstacle). The algorithm does not change its behavior (i.e., circumventing the obstacle each time results in the same strengthening of links), which implies that it does not learn (because all learning involves a change in behavior). The robot, however, *does* learn: the way it responds to the proximity of the obstacle changes as the result of a regularity in its environment. That said, there may be other cases where the learning of a robot is conditional on the learning of an algorithm. Assume that the extent to which the algorithm changes the strength of a link reduces each time the robot manages to overcome the obstacle (i.e., that the change in link strength is bigger after the first successful trial than after the second one, that it is bigger after the second trial than after the third one, and so). In this case, the algorithm does learn: its response to overcoming the obstacle changes because of a regularity (i.e., the repeated experience of overcoming an obstacle). These examples illustrate that conclusions about whether a particular system (e.g., a humanoid robot, algorithm) learns, depend only on an intrasystem analysis (i.e., on whether there are arguments to say that a regularity changes the way in which a system behaves towards a specific stimulus). Intersystem analyses come into play only when considering how different systems interact. [4]

As a second example, let us return to genetic sensitization (see Figure 4). Imagine that the low transcription levels of the stress response genes in response to dehydration are due to the fact that the genes are covered with a substance that reduces the rate of transcription (e.g., methyl). When dehydration is experienced, an epigenetic (e.g., demethylation) mechanism becomes active that removes part of this material from the genes, thus allowing for a higher level of transcription in response to future instances of dehydration (see Wu & Zhang, 2010, for evidence for such a mechanism). Within such a scenario, the stress response genes have learned (i.e., altered their activity in response to dehydration as result of the repeated experience of dehydration) in a way that is conditional upon the behavior of an epigenetic mechanism (i.e., removal of methyl in response to dehydration). If the behavior of the epigenetic mechanism towards dehydration also changes as the result of the fact that dehydration occurs repeatedly (e.g., it removes more methyl in response to dehydration the more often dehydration is experienced), then the epigenetic mechanism can also be said to have learned. Also note that the learning of the stress response genes could in turn be a condition for learning at the level of cells and the whole organism (e.g., as the result of the

[4] Such interactions can be looked at from both a functional perspective and a mechanistic perspective. From a functional perspective, the learning of one system can be said to be conditional on the behavior or learning of another system. This merely implies that learning in the first system (i.e., the impact of regularities in the environment of that system on how the system responds to a particular stimulus in its environment) is itself a function of the behavior or learning of a second system (i.e., how the second system responds to a stimulus or how regularities change the way in which the second system responds to a stimulus). From a mechanistic perspective, the learning of one system can be said to be mediated by the behavior or learning of another system. This implies that the behavior or learning of the second system is a necessary step via which regularities in the environment change the behavior of the first system. Hence, unlike a functional perspective on intersystem analyses, a mechanistic perspective on intersystem analyses implies contiguous causation, that is, the idea of a chain of events in which one event puts into motion another event (see Chiesa, 1992, De Houwer & Hughes, 2020, and Hayes and Brownstein, 1986, for a more detailed discussion on the distinction between functional and mechanistic approaches in science). Note that the extended functional definition of learning highlights that the parts of a mechanism are not necessarily static but can behave and learn.

repeated experience of dehydration, the retention of water by cells and the whole plant in response to dehydration increases). This illustrates that the behavior and learning of more than two systems can be interrelated. [5]

As a final example, consider the school of fish that changes its response to the shipwreck (Figure 6). It might be the case that the learning of the school is conditional on the fact that individual fish in the school learn to stay away from shipwreck. According to our definition, an individual fish has learned if its behavior towards a stimulus changes (e.g., changes its response to the proximity of the shipwreck) because of regularities in its environment (e.g., a predator attack when it approaches the shipwreck). Other fish in the school, however, might display the same transitions in state without having learned anything. For instance, it could be that some fish in the school always copy the transitions in state of the fish that surround them (Herbert-Read et al., 2015). Hence, if some fish learn to stay away from the shipwreck, these other fish also stop approaching the shipwreck. The latter fish, however, have not learned but merely continue to behave in the same way: they always mimic the transition in state of the fish surrounding them. Nevertheless, the group itself can be said to have learned, and this learning was conditional upon learning at the level of certain individuals.

Although these examples illustrate that the extended functional definition allows for intersystem analyses, please keep in mind that it does not require this type of analyses. Even when it is not possible to relate learning and behavior in one system to learning and behavior in other systems, there is still merit in analyzing learning and behavior in the first system.

---

[5] This idea resembles the idea of multilevel-selection theory in biology (Okasha, 2006). One important difference is that evolutionary selection involves successive generations of systems (e.g., a particular species of animal) whereas learning involves individual systems (e.g., a particular member of an animal species). Another difference is that selection refers exclusively to changes as the result of consequences whereas learning refers to changes due to consequences of behavior (i.e., operant conditioning) but also changes due to regularities in the presence of one or multiple stimuli (e.g., habituation, classical conditioning).

Consider the example of deep learning in artificial neural networks (LeCun et al., 2015). With deep learning networks, it is often difficult to comprehend which changes in the networks produce changes in the output of the networks (i.e., the network can be conceived of as a black box). Also in these cases, intrasystem analyses allow one to analyze the learning and behavior of the network. More specifically, by revealing the environmental events that control the behavior and learning of the network, one can predict and influence the output of the network solely on the basis of intrasystem analyses. This can be achieved by observing and manipulating the environmental events that the intrasystem analyses identifies as controlling the output of the system. In sum, there is merit in intrasystem analyses even in the absence of intersystem analyses.

Clarifying the Boundaries of Behavior and Learning: Increased Conceptual Precision

The final example of the previous section not only illustrates the complex relations that can exist between learning and behavior of different systems but also highlights that, because of the functional nature of our definitions, claims about behavior and learning always involve causal assumptions. For instance, the same state transition (e.g., from moving forward to turning left) can be labelled as different behaviors (e.g., mimicking other fish vs. avoiding a shipwreck) depending on what the state transition is a function of (e.g., the state of other fish vs. the proximity to a shipwreck). Likewise, whether learning has occurred cannot simply be observed by looking at changes in the state of systems. It also requires two causal assumptions. First, learning requires a change in behavior towards a stimulus, that is, a change in the manner in which this stimulus influences the state of a system at different moments in time. Hence, the claim the learning has occurred depends on assumptions about the stimulus that influences the state of a system at different moments in time. Second, a change in behavior towards a stimulus qualifies as learning only if it is due to a regularity in environment (e.g., the repeated presence of a stimulus, the pairing of stimuli, or the fact that a behavior results in a stimulus). The importance of causal assumptions when applying our definitions of learning and behavior is a direct consequence of the functional nature of these definitions: they imply that the state of system (in the case of claims about behavior) or the behavior of a system (in the case of learning) is a function of specific events in the environment (Hayes & Brownstein, 1986). Hence, behavior is more than a state transition (i.e., the state transition should be a function of a specific stimulus), a change in behavior is more than a change in state transitions (i.e., the change in state transitions should reflect a change in the way that stimulus influences the state of the system), and learning is more than a change in behavior towards a stimulus (i.e., the change in behavior should be a function of a regularity). These distinctions allow researchers to communicate in precise ways and thus avoid confusion.

Because learning requires not only the observation of a change in state transitions but also two causal assumptions, we can delineate learning from other phenomena. First, there can be no learning when there is no change in state transition. Imagine that during manufacturing, a programmer programs the neural net of a humanoid robot in such a way that the robot always turns left when it nears the position of the obstacle, so that it always circumvents the obstacle. In that case, the robot has not learned to circumvent the obstacle but has been programmed to behave in a fixed way (turn left when nearing the obstacle). There is no *change* in the robot's behavior (i.e., from the very start, the robot turns left in proximity of the obstacle) so it does not make sense to say that the robot has learned to overcome the obstacle. Likewise, in living organisms, there are some behaviors that are present from birth and that could be seen as "programmed" by evolution (e.g., the grip reflex in babies). In these cases, one would say that evolution functioned as a programmer for those behaviors.

Second, some phenomena do involve changes in state transitions but do not qualify as instances of learning because the changes in state transitions are unlikely to be changes in behavior towards a specific stimulus. Consider the effects of damaging a system. In line with what we noted at the start of this section, a change in behavior towards a stimulus implies that the impact of that stimulus on the state of a system has changed from Time 1 to Time 2. If, however, the system is damaged between Time 1 and Time 2, then a comparison of the state transitions at those two points in time no longer informs us about changes in the impact of the stimulus. This is because the damage inflicted on the system is an alternative cause of the change in state transitions. Imagine a humanoid robot that moves forward, bumps into an obstacle, then moves backward, moves forward again, bumps into the obstacle again, and so on. After many repetitions of these events, the robot suddenly stops moving forward before it reaches the obstacle. One analysis of these events is that the robot has learned: the way in which the robot behaves towards the obstacle has changed (i.e., initially, the presence of the obstacle has no impact on the robot; eventually, the presence of the robot causes it to stop moving forward) because of a regularity in the environment (e.g., moving forward when nearing the obstacle results in bumping against the obstacle). Another possible analysis, however, is that the robot eventually stops moving forward because it has become damaged in such a way that it is no longer capable of moving forward. In this case, the fact that the robot eventually stops rather than continues to move forward is not due to a change in the impact of the obstacle on the robot (i.e., the change in state transition would have occurred even if the robot remained irresponsive to the obstacle) but simply to the damage inflicted on the robot. Hence, it does not make sense to say that the way in which the robot responds to the obstacle has changed, which implies that it also does not make sense to say that the robot has learned. Note that this reasoning applies also to temporary damage such as a temporary depletion of energy (e.g., fatigue, depleted batteries). Also in this case, changes in state transition do not inform us about changes in the impact of a specific stimulus on the state of a system. Also in this case, the system does not change its behavior towards a specific stimulus but merely becomes incapable of responding to stimuli, be it only for a limited period in time (e.g., until the batteries are replenished).

Third, some changes in state transitions do qualify as changes in behavior towards a specific stimulus but do not qualify as learning because they are not due to regularities in the environment. One example might be the disappearance of the grasp reflex in babies (see De Houwer & Hughes, 2020, p. 5). During the first months of their lives, babies will close their hand when an object touches the palm of the hand, thereby grasping the object (e.g., a finger). This behavior disappears after 5 or 6 months. Hence, there is a change in behavior: the same stimulus (e.g., a finger touching the palm of the hand) leads to a different transition in the state of the baby at birth (i.e., from open hand to closed hand) than when the baby is 6 months old (i.e., the hand remains open). Whether this change in behavior qualifies as an instance of learning depends on its causes. If, for instance, it occurs only if the baby experiences a certain regularity in its environment (e.g., repeated stimulation of the hand), then the change in responding to the stimulation of the hand would qualify as an instance of learning (more specifically, habituation). If, however, it occurs independently of regularities in the environment of the baby, then it would not qualify as an instance of learning according to the (extended) functional definition of learning (also see De Houwer & Hughes, 2020, p. 5). For instance, it might be due to maturation, that is, to the spontaneous development of the central nervous system. In this case, one could say that not only the grasp reflex but also the disappearance of the grasp reflex has been programmed by evolution. An analog of this in robots would be that during manufacturing, the neural net of the robot has been programmed to circumvent an obstacle but also to stop doing so after a certain period of time.

Part IV: Limitations

Difficulties in Establishing that Learning has Taken Place

As we previously noted, the application of the extended functional definition of learning requires causal assumptions. At the conceptual level, this has the advantage that clear distinctions can be made between learning and other phenomena. At the empirical level, however, it is difficult to verify that learning has occurred simply because causality cannot be observed directly but can only be inferred. This is a limitation of all functional definitions (also see De Houwer et al., 2013), including the extended functional definition of learning. Fortunately, the elements that are causally related in the extended functional definition (i.e., the state of a system, state transitions, changes in state transitions, stimuli, and regularities) can often be observed or manipulated directly by the researcher who analyses the system. To verify causal assumptions, one can thus manipulate the presence and nature of stimuli and regularities and observe whether this results in transitions in states and changes in state transitions. Given appropriate controls, this allows one to infer instances of behavior, changes in behavior, and learning. One should be aware that without such experimental evidence, statements about behavior, changes in behavior, and learning always remain hypotheses. Also, whereas the extended functional definition provides conceptual tools that can facilitate debates about learning in various systems, using the definition by no means guarantees that different researchers will always agree about whether learning has occurred in a particular system.

Although this is a significant limitation of the extended function definition of learning, it is important to note that also other definitions of learning are difficult to verify. For instance, as Barron et al. (2015) noted, many definitions of learning refer to the storage of information. These definitions have the downside that (the storage of) information cannot be observed directly given the nonphysical nature of information (e.g., Wiener, 1961). One could hope to find observable proxies of the storage of information but this requires (1) the assumption that there is a one-to-one correspondence between an observable proxy and the storage of information (i.e. that the proxy is influenced by and only by the storage of information), (2) assumptions about which events carry information and which information carried by those events is stored, (3) assumptions about when and how information that is stored influences observable behavior. Hence, arguably such definitions require even more causal assumptions that are even more difficult to verify.

Despite acknowledging this downside of definitions of learning that refer to the storage of information, Barron et al. (2015) still preferred these definitions to functional definitions in terms of changes in behavior because, in their opinion, functional definitions perform more poorly in terms of delineating learning from other phenomena. As we argued above, this criticism does not apply to the extended functional definition of learning (see the section on clarifying the boundaries of behavior and learning). Moreover, as we argued throughout the paper, the extended functional definition of learning can be used to study learning in different systems. Finally, because it allows for intersystem analyses, it allows researchers to relate the behavior and learning of multiple systems. These intersystem analyses have the unique advantage that they deploy the same concepts of learning and behavior within each of the systems that are related.

In sum, although the extended functional definition has the downside that it is difficult to verify that learning has occurred, arguably other definitions are more problematic in this and other respects.

Deviations from Everyday Language

Together with other definitions that define learning in terms of changes in behavior, the extended functional definition of learning deviates from how the term “learning” is typically used in everyday language. As noted earlier, for many lay people and scientists alike, learning is some kind of mechanism for storing knowledge. From this perspective, learning is not a subclass of changes in behavior (e.g., those changes that are due to regularities in the environment) but a mechanistic explanation for changes in behavior (e.g., the formation of informational representations that can determine behavior). Many concepts in science, however, are used in ways that deviate from their everyday use (e.g., gravity as a curvature in spacetime rather than an invisible force). In many cases (as in the case of gravity), scientific progress depended on letting go of the everyday meaning of concepts in science. Time will tell whether the extended functional definition will foster scientific progress. For now, it is important to note that this is its only aim. We do not want to claim that the extended functional definition is the “true” definition of learning. We only claim that adopting it can have advantages for the study of learning in various scientific disciplines.

Readers for whom a non-mechanistic, functional definition of learning sits uncomfortable could opt to use the term “learned behavior” when referring to learning as a phenomenon and maintain the term “learning” for referring to the mechanisms that mediate learned behavior. Our extended functional definition could then be used as a definition of learned behavior, which would keep intact all the benefits of using the extended functional definition across different scientific disciplines. Having said this, we believe that such a broad conceptualization of “learning” would have little added value. Because it is likely that learning mechanisms vary widely within and across systems, a broad mechanistic conceptualization of “learning” would imply little more than the idea that learned behavior must be due to some kind of mechanism. An alternative would be to restrict learning to only a subset of mechanisms that can produce learned behavior but we fear that this will result in unproductive debates about what are the “true” learning mechanisms. Also, because it is by definition at least as difficult to establish the presence of a mechanism that produces learned behavior than it is to establish the presence of learned behavior, we fear that restricting learning to a subset of mechanisms would hamper the study of learning. Hence, we believe that scientific progress is served best by using the term “learning” in a non-mechanistic sense as is the case with the extended functional definition of learning that we put forward in this paper.

We want to avoid ontological debates not only about the nature of learning but also about the nature of systems, states, and behavior. Using our definitions, one might well construct analyses that do not sit well with pre-scientific intuitions. For instance, some might have difficulties with the idea that inanimate objects like rocks and planets might behave. Although we hope that our definitions lead to interesting debates about the nature of systems, states, behavior, and learning, we do not wish to become trapped in such ontological debates. Our main hope and ambition is that the definitions put forward in this paper will help scientists achieve their scientific goals.

Conclusion

Definitions are tools at the service of better science. Just as it is good to reflect about the utility of other tools, it can be good to reflect on the utility of definitions (Machado & Silva, 2007). In this paper, we introduced and debated a new definition of learning that, in our opinion, has high utility because it is at the same time precise and broadly applicable. We hope that, because of these features, the extended functional definition of learning will facilitate communication between researchers from various scientific disciplines and stimulate new research on learning both within and across disciplines.

References

Avramova, Z. (2015). Transcriptional 'memory' of a stress: transient chromatin and memory (epigenetic) marks at stress-response genes. *The Plant Journal, 83(1),* 149–159. https://doi.org/10.1111/tpj.12832

Barron, A. B., Hebets, E. A., Cleland, T. A., Fitzpatrick, C. L., Hauber, M. E., & Stevens, J. R. (2015). Embracing multiple definitions of learning. *Trends in Neurosciences, 387*, 405-407.

Bouton, M. E. (2016). *Learning and behavior: A contemporary synthesis* (2nd ed.). Sinauer.

Burgos, J. E. (2018). Is a nervous system necessary for learning? *Perspectives on Behavior Science*, *41*, 343-368.

Catania, A. C. (2013). *Learning.* Cornwall-on-Hudson, NY: Sloan.

Chiesa, M. (1992). Radical behaviorism and scientific frameworks: From mechanistic to relational accounts. *American Psychologist*, *47*, 1287–1299.

Chittka, A., & Chittka, L. (2010). Epigenetics of royalty. *PLoS Biology*, *8(11)*, e1000532. https://doi.org/10.1371/journal.pbio.1000532

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.

De Houwer, J., & Hughes, S. (2020). *The psychology of learning: A functional-cognitive introduction.* The MIT Press.

Ding, Y., Fromm, M. & Avramova, Z. (2012). Multiple exposures to drought 'train' transcriptional responses in Arabidopsis. Nature Communications, 3, 740, https://doi.org/10.1038/ncomms1732

Freeberg, T. M. (2004). Social transmission of courtship behavior and mating preferences in brown-headed cowbirds, Molothrus ater. *Learning & Behavior, 2004,* 32 (1), 122-13.

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., ... & Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research, 17(6),* 669-681.

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *Behavior Analyst, 9*, 175–190.

Herbert-Read, J. E., Buhl, J., Hu, F., Ward, A. J. W., & Sumpter, D. J. T. (2015). Initiation and spread of escape waves within animal groups. *Royal Society Open Science*, 2140355140355. http://doi.org/10.1098/rsos.140355

Hughes, S., & De Houwer, J. (2020). Genetic learning: A new conceptual framework for studying the impact of the environment on changes in genetic behavior. Unpublished manuscript. DOI: 10.13140/RG.2.2.33044.63364

Kucharski, R., Maleszka, J., Foret, S., & Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science, 319(5871)*, 1827–1830. https://doi.org/10.1126/science.1153069

Lachman, S. J. (1997). Learning is a process: Toward an improved definition of learning. *Journal of Psychology, 131*, 477-480.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444.

Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. American Psychologist, 62, 671–681.

Moore, D. S. (2015). *The developing genome: An introduction to behavioral epigenetics*. Oxford University Press.

Okasha, S. (2006). *Evolution and the levels of selection*. Oxford University Press.

Pietraszewski, D. (2021). Towards a computational theory of social groups: A finite set of primitives for representing any and all social groups in the context of conflict. *Behavioral and Brain Sciences,* 1-62. doi:10.1017/S0140525X21000583

Rahwan, I., Cebrian, M., Obradovich, N. et al. (2019). Machine behaviour. *Nature, 568*, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Thompson, R. F. (2009). Habituation: A history. *Neurobiology of Learning Memory, 92*, 127–134.

Virlouvet, L., & Fromm, M. (2014). Physiological and transcriptional memory in guard cells during repetitive dehydration stress. *New Phytologist, 205 (2)*, 596-607.

Wiener, N. (1961). *Cybernetics, or control and communication in animal and machine* (2nd Ed.). The MIT Press.

Wu, S. C., & Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nature Reviews: Molecular Cell Biology, 11(9)*, 607-620.