The pitfall of neurobiological reductionism

Pierre Perruchet and Annie Vinter *
Université de Bourgogne, LEAD-CNRS, Dijon, France

ABSTRACT

Although we subscribe to the idea of promoting associationism, the project of elaborating a general model relying only on neurobiological data seems doomed to failure. For several decades now, in departure from the conceptions of Pavlov, Thorndike or Skinner that served as references for Hebb, first-hand workers on associative learning have considered that associations take place between *mental* representations, possibly *complex* ones. The laws governing their formation and evolution apply at a level of explanation other than the biological one. The shortcomings of neurobiological reductionism are not due to the fact that knowledge in this field is still incomplete, but to the need for considering the mental level as causal in cognitive sciences. We suggest that the field of dynamical systems, involving the related concepts of emergence, reciprocal causality, and self-organization, provides the best framework to conceive the across-time interplay between mental, biological, and environmental events.

Keywords: Associative learning; reductionism; dynamic systems; mental states

Le piège du réductionnisme neurobiologique

RÉSUMÉ

Bien qu'étant favorables à l'idée de promouvoir l'associationnisme, le projet de donner corps à un modèle général en se fondant uniquement sur les données neurobiologiques nous semble voué à l'échec. Depuis plusieurs décennies, loin des conceptions initiales de Pavlov, Thorndike, ou Skinner servant de références à Hebb, les chercheurs centrés sur l'apprentissage associatif considèrent que les associations s'opèrent entre représentations *mentales*, éventuellement *complexes*. Les lois qui régissent leur formation et leur évolution s'appliquent à un autre niveau d'explication que le niveau biologique. Les limites d'un réductionnisme neurobiologique ne sont pas liées au caractère encore incomplet des connaissances en ce domaine, mais à l'obligation d'intégrer les états mentaux dans les sciences cognitives. Nous

* Corresponding author : Pierre Perruchet, LEAD, Esplanade Erasme, Université de Bourgogne - 21000 Dijon pierre.perruchet@u-bourgogne.fr.

proposons que le champ des systèmes dynamiques, impliquant les concepts connexes d'émergence, de causalité réciproque, et d'autoorganisation, fournit le meilleur cadre théorique pour concevoir les interactions continues entre événements mentaux, biologiques, et leur contexte environnemental.

Mots-clés : Apprentissage associatif ; réductionnisme ; systèmes dynamiques ; états mentaux

Arnaud Rey should first be congratulated for proposing a metatheoretical framework, a kind of contribution that unfortunately seems to be on the way to extinction in contemporary psychology. We wholeheartedly endorse the radical tone of the proposal, for the same reasons as those put forward by the author: Radical statements are easier to falsify, and therefore more heuristic, than positions admitting both a thing and its opposite. Also, the choice of the concept of associations as a cornerstone in this context seems to be a good starting point, because of its apparent simplicity and its direct openness to the crucial notion of learning.

However, as we progressed through the first few pages of the paper, we felt a growing sense of disappointment, culminating in a profound disagreement with the proposal to pursue the Hebbian project, and to reshape the ill-defined psychological notions "with more precise concepts rooted in the field of neurobiology".

There are two components in Hebb (1949)'s famous book. The best-known is his law, quoted in the article, of how an association can be encoded in the brain. Remember that Hebb was a psychologist whose law is nothing more than a hypothesis about how the brain can form memories. Admittedly, data consistent with this hypothesis were later observed, initially at the level of the hippocampus, and the mechanism is now studied under the name of Long-Term Potentiation (LTP). But by the same token, it became clear that this was not the only mechanism explaining brain plasticity. For example, the converse phenomenon, Long-Term Depression (LTD), was also observed in the hippocampus before being found in other brain regions (e.g., Hansel & Bear 2008 for a review). We guess that the fascination that Hebb's law seems to exert today on some contemporary researchers in neural computing is motivated more by its suitability for use in computer simulations than by neurobiological evidence.

But this law is only an instantiation of Hebb's more general project of grounding the explanation of mental events in neurobiology. It is this more general proposal, endorsed in Arnaud Rey's article, with which we fundamentally disagree.

Let us start with an observation that can hardly be disputed. Over the past 75 years since Hebb's book, the study of associative learning, whether coined as conditioning, statistical learning or something else, has led to the discovery of a considerable number of important phenomena, including those mentioned by Arnaud Rey (blocking, overshadowing…). Yet *none* of these advances has been guided by the Hebbian project, and more generally by neurobiological concerns.

Historically, the neuron is *not* at the basis of associationism, contrary to Arnaud Rey's assertion. The paper's subtitle "philosophical and neurobiological associationisms" simply overlooks the main territory of associationism, namely psychology.

Within the field of psychology, the ubiquitous reference to the works of Pavlov, Thorndike or Skinner in most psychology textbooks is potentially misguiding. As has long been acknowledged, the concepts they introduced are underpowered to explain all human behavior. The "official doctrine", so to speak, of contemporary first-hand workers in associative learning is that associations occur between *mental representations*. The following quotation, borrowed from one of the leading theorists of animal learning, expands on this notion:

> "Properly understood associative learning theory is remarkably powerful. Of course, such a theory must reject the restrictive assumption of S-R theory, which allowed associations to be formed only between a stimulus and a response, and should assume that a representation of any event, be it an external stimulus or an action, can be associated with the representation of any other event, whether another external stimulus, a reinforcer, the affective reaction elicited by the reinforcer, or an animal's own actions. Equally important, however, it must allow that the representation of external events that can enter into such associations may be quite complex. They need not be confined to a faithful copy of an elementary sensation such as a patch of red light; they may be representations of combinations or configurations of such elementary stimuli; they may even include information about certain relationships between elementary stimuli. But once we have allowed associative learning theory these new assumptions, we have a powerful account, capable of explaining quite complex behavior – including behavior that many have been happy to label cognitive and to attribute to processes assumed to lie beyond the scope of any theory of learning." *(*Mackintosh 1997, pp. 883-884*)*

Replacing the elementary stimuli of archetypal conditioning settings with complex mental representations obviously moves away from the

belief that simplistic biological mechanisms are sufficient to provide a satisfactory account of associationism. Is there any chance that advances in neurobiological research will change this state of affairs? Our response is negative. We contend that neurobiology alone will *never* explain behavior and mental states, whatever the time horizon, for a principled reason outlined below.

Let us take the example of an ant colony, whose behavior would seem a priori to be well suited to the biological reductionism advocated by Arnaud Rey. It appears that, among other remarkable organizational features, the ants use the shortest paths to join their anthill to the food sources. Each ant displays two potentially relevant characteristics, namely the propensity to follow a trail of pheromones and to strengthen this trail by depositing on it the same chemical substance. However, those characteristics by themselves do not explain the selection of the shortest path by the colony. The now well-known explanation is as follows. The ants first randomly forage for food and come back to the anthill when they have discovered a food source. So doing, some fortunate ants naturally find the food through a more direct path than others. The ants that have discovered the shortest path will cover the back and forth distance between the anthill and the food source fastest, and hence, more often than other ants in the same amount of time. Because the ants deposit pheromones throughout their walk, the accumulation of pheromones on the path of the lucky ants is fastest, hence progressively attracting the other ants (e.g., Camazine et al., 2003).

What role does biology play in this explanation? Undoubtedly, the biological processes underlying the propensity to follow and strengthen a pheromone trail are a part of the explanation.

But to account for the colony's selection of the shortest path, we need to move away from the biological level. We need to consider the whole system, including other ants and the environment, from a dynamical perspective (i.e. in considering the *across time* interactions of the system components). For instance, the explanation relies on the fact that a shorter path is covered more often than a longer path in a given period. It also relies on a chemical property of the pheromones (their diffusion through a gradient, see e.g., Kugler & Turvey, 1987). Moreover, the explanation calls for concepts such as emergence (a new collective behavior emerges from the summation of individual behaviors), backward causation (the emergent discovery of the shorter path influences in turn subsequent biological processes) and self-organization (because there's no need to call for an external planning agent).

If ant behavior can't be explained by biological processes alone, it seems likely that human behavior can't either. We contend that human mental activities are not reducible to their biological underpinnings, whatever the advances in our biological knowledge, because mental activities can be understood only in a framework involving multiple levels of causality, such as in the field of dynamical systems (e.g., Smith, 2005). Providing a full-blown explanation of how dynamical models can account for the formation of complex mental representations is beyond the scope of this comment (see Perruchet & Vinter, 2002). However, since Arnaud Rey's paper argues in favor of neurobiological reductionism, a few words about the concept of reciprocal causality are in order. The concept of reciprocal causation (still coined as upwards-downwards, double, or circular causality) allows to account for the prima facie conscious experience of mental causation without violating the causal closure of the physical domain (e.g., Varela & Thompson, 2003). In a nutshell, the mental level is conceived as an emergent property of neural processes. This means that mental activities are realized in the brain, but are not reducible to biological events, just as the selection of the shorter path by the ant colony is due to the perception and production of pheromones but is not reducible to these biological processes. Mental states are endowed with their own organizational rules, which makes them causal when inserted in a dynamic interplay with biological mechanisms.

In his plea for a neurobiological reductionism, Arnaud Rey criticizes psychology for using ill-defined concepts. We can only agree on the fact that most psychological concepts lack clear, consensual definitions, including those used in the dynamical framework. For instance, the pivotal concept of emergence turns out to be defined by different criteria in different fields of research. This raises two questions.

The first is: Is there any advantage in couching ill-defined concepts in neurobiological terms? Let's replace "attention" with "process", as Arnaud Rey suggests, or with allegedly "more precise" terms such as assembly of neurons, or still "the cascade of activated sub-assemblies of neurons". Since all mental states involve biological processes, this would lead to using the same terminology to designate all classical psychological concepts, such as perception, memory, reasoning, emotion, motivation, language, consciousness and so on. The add-on precision is questionable, to say the least. Psychological concepts have at least one advantage: To generalize William James' famous formula on attention, which Arnaud Rey evokes, everyone understands what they are dealing with.

The second question raised by the lack of a consensual definition of psychological concepts is whether this really hampers the advancement of

knowledge. We would be more inclined to argue the opposite. When the term artificial intelligence (AI) emerged in the 1950s, its main component, intelligence, was far to receive a consensual definition. This indeterminacy could have fueled endless debates about whether intelligence is binary or continuous, unidimensional or multidimensional, and so on. Instead, as Melanie Mitchell (2019) notes: "For better or worse, the field of AI has largely ignored these various distinctions". Interestingly, she adds that in a 2016 report on the current state of AI, a committee of prominent researchers pointed out that "the lack of a precise, universally accepted definition of AI probably has helped the field to grow, blossom, and advance at an ever-accelerating pace." (p. 20). We surmise that those who now contend that psychological concepts are too elusive to deserve consideration would have, sixty years ago, been left wondering what intelligence means exactly, without taking a step forward.

In conclusion, promoting a radical associationism seems an interesting option. For it to have a chance of success, however, it seems necessary to consider that associations involve complex mental representations, the formation of which requires the full power of dynamic, selforganizing systems. The Hebbian notion of neural assemblies seems notoriously underpowered for this endeavor. Talking about assemblies of neurons is all the less appealing given that nobody knows how electrical and biochemical phenomena generate mental states and consciousness. This ignorance is problematic if we're trying to explain everything on a neurobiological level, but much less so in a psychological approach that acknowledges multiple levels of explanation. Nobody knows why two masses, even if separated by light-years in what appears to be a vacuum of matter, attract each other. The physical substrate of this force, or even whether it is really a force comparable to the other three fundamental forces (electromagnetic, weak and strong nuclear interactions), remains an open question. Yet this did not prevent Newton from establishing the law of universal gravitation, according to which every particle attracts every other particle in the universe with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between their centers. Likewise, psychology should not refrain from using concepts that are deprived of known biological correlates, even though the search for these correlates is obviously a worthwhile and promising objective.

REFERENCES

Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. J. (2003). *Self-Organization in Biological systems.* Princeton : Princeton University Press.

Hansel, C., & Bear, M.F. (2008). LTD- synaptic depression and memory storage. In Molecular mechanisms of memory. Vol.4 of *Learning and memory: A comprehensive reference* (J. Byrne, Ed.) (pp. 327-365). Oxford : Elsevier.

Hebb, D.O. (1949). *The organization of behavior: A neuropsychological theory.* Hoboken : John Wiley & Sons, inc.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement.* Hillsdale : Erlbaum Associates.

Mackintosh, N. J. (1997). Has the wheel turned full circle? Fifty years of learning theory, 1946-1996. *The Quarterly Journal of Experimental Psychology*, 50, 879-898.

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans.* New York : Farrar, Straus & Giroux.

Perruchet, P., & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences, 25*(3), 297-388.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review*, 25, 278-298.

Varela, F,J., & Thompson, E. (2003). Neural synchrony and the unity of mind: a neurophenomenological perspective. In A. Cleeremans (ed.). *The unity of consciousness: Binding, integration, and dissociation*. Oxford: Oxford University Press (pp. 266-287).