Embracing integration and complexity: placing emotion within a science of brain and behaviour

Luiz Pessoa

Department of Psychology and Maryland Neuroimaging Center, University of Maryland, College Park, MD, USA

Abstract

The present paper addresses conceptual issues that are central to emotion research. What is emotion? What are its defining characteristics? The field struggles with questions like these almost constantly. I argue that definitions, and deciding what is the proper status of emotion, are not a requirement for scientific progress - in fact, they can hinder it. Therefore, "emotion" researchers should strive to develop a science of complex behaviours, and worry less about their exact nature. But for interesting behaviours, is most of the explaining that is needed present at the level of isolated systems (perception, cognition, etc.) or at the level of interactions between them? I suggest that the level of interactions is where most of the work is needed. Accordingly, I advocate that it is important to embrace integration, and not to strive to necessarily disentangle the multiple contributions underlying behaviours. More generally, it is argued that we need to revise models of causation adopted when reasoning about the mind and brain. Instead, a "complex systems" approach is required where the interactions between multiple components lead to system-level - emergent - properties that cannot be isolated or attributed to more elementary parts.

Stop worrying about the definition of emotion

Unlike other areas of mind-brain investigation, emotion research has been constantly in a state of existential self-examination. What is emotion? What are its defining characteristics? Are emotions distinct from feelings? The field has struggled with such questions for a long time - and continues to grapple with them. For example, in the just-published edition of the authoritative *The Nature of Emotion* (Fox, Lapate, Davidson, & Shackman, 2018), "What is an emotion?" is still a question that is addressed across nine essays.

Naturally, there is value in developing a "scientifically useful theory of emotions" (Adolphs & Andler, 2018). In terms of the brain, a neuroscientist may be interested in studying brain

structures such as the amygdala or the periaqueductal grey in a set of tasks that are very well defined and controlled. In terms of behaviour, a behavioural researcher may choose to investigate simple forms of aversive classical conditioning. It is conceivable that under such circumstances one could label the object of their study as "emotion." Both in terms of brain and behaviour, it is also conceivable that a larger set of conditions could be addressed and coherently labelled "emotion" (Figure 1A).

But definitions, and deciding what is the proper status of emotion, are not a requirement for scientific progress. In fact, definitions can even hinder growth if the conceptual boundaries prove to be stifling. For example, had physicists decided that it was imperative to establish if light functions as a particle or a wave, it would have been almost certainly detrimental to progress; for one, the wave-particle duality is at the core of the foundation of quantum mechanics. Consider another example, again from physics. According to the so-called Standard Model of particle physics, three forces explain phenomena at the scale of "small objects." The model does not include gravitation, which obviously matters for large objects. Whereas particle physics and gravitation could be considered separate areas of research, the goal to integrate all four fundamental forces in more complete models has constituted an active field of investigation. And, in this pursuit, physicists are not shunned because particle physics (for the sake of comparison, consider emotion) is different or separate from gravitation (consider cognition).

A specific goal of mind-brain research is to arrive at a scientifically useful theory of emotions. However, a broader goal, which I suggest is more important, is to develop a science "to explain the complex behaviors of people and animals" (Adolphs & Andler, 2018). My claim in the present paper is that complex behaviours do not obey the boundaries of research areas - such as perception, cognition, action, emotion, and motivation - in any useful way.

For most mammals, and indeed many vertebrates, complex behaviours are common, and considerably more frequent and sophisticated than realised even fairly recently (for example, Clayton & Emery, 2015; Rischawy, Blum, & Schuster, 2015) - their world is not the constrained setting of laboratory experiments. The question that needs to be considered then is as follows: For "interesting behaviors" that we care about (see the last section), is most of the explaining needed to be done at the level of isolated systems (perception, cognition, etc.) or at the level of *interactions* (see the next section) between them? To help in the discussion, consider the view articulated by Adolphs and Andler (2018) expressing their interest in a level of behavioural complexity situated between reflexes and volitional behaviour, which they call the level of "semi-flexibility," and according to them the rightful place of emotion.

But does this level capture the behaviours of interest to an emotion researcher? Whereas it is conceivable that it does, let's not close doors, and instead allow questions to be investigated empirically without prejudgment ("oh, but that is cognition, and of course emotion interacts with cognition"). Here, briefly, it is important to consider that "volitional behavior" is also characterised by properties of "semi-flexibility." A growing number of investigators have persuasively argued that "control" is much more "automatic" than previously understood (see Anderson, 2017; Moors, Boddez, & De Houwer, 2017). For example, Moors et al.

(2017) discuss the idea that *goal*-directed processes should *not* be viewed as restricted to emotion regulation but also play a role in the initial causation of emotional action tendencies. They suggest that, if emotional actions are defined as ensuing from action tendencies with control precedence (those that take priority over other action tendencies), because these action tendencies ensue from stimuli that are relevant for highly valued goals, then there should be no a priori reason to deny the role of goal-directed processes in the *causation* of emotional actions. More generally, several investigators have proposed that the distinction between goal-directed and stimulus-driven information processing needs to be reconsidered (Anderson, 2017; Awh, Belopolsky, & Theeuwes, 2012; Melnikoff & Bargh, 2018; Pessoa, 2013). In a complementary fashion, there is a considerable body of research documenting the ways in which "automatic" emotional processing is influenced by awareness, attention, context, and other high-level factors (for example, Pessoa, 2005; Pessoa, 2013).

My goal here is not to review the extensive literature informing the issues above but to highlight the difficulty of carving behaviours and phenomena that fall within the "rightful place of emotion" (Figure 1B). Ultimately, one of the implications of the present paper is that this place simply does not exist.

Embracing integration

Models of the mind and brain are peppered with diagrams with boxes and arrows that link purported mental functions to each other, and/or to brain regions. Given that the mind-brain sciences have been doing this for 150 years (see Shallice, 1988), the language utilised by scientists to describe mental and neural processes is, by and large, fairly modular: "perception," "attention," "motivation," "cognition," and so on.

As an illustration of the conceptual issues investigators face, let's examine a previous "attentional blink" experiment from my lab, where participants watched a stream of rapid, flashed images containing among them either a house or a skyscraper (Lim, Padmala, & Pessoa, 2009). Participants had to identify which of these scenes was present in the video. Before viewing the clips, half of the participants received a mild electric shock while viewing a series of skyscrapers; the other half, instead, watched a series of houses appear, paired with the same mild shock. We found that participants conditioned to the skyscrapers detected them better than houses; conversely, participants conditioned to houses detected them better than skyscrapers. And in each case, responses in the visual cortex were stronger for the type of stimulus (house or skyscraper) to which participants had been conditioned. This study shows that perception is not a passive process that merely reflects the external world. Rather, it involves picking up on the significance of objects, which determines how they are processed.

This type of phenomenon, which can be called *affective attention*, has been extensively studied in the literature (Pessoa, 2005; Vuilleumier, 2005). Conceptually, should we view it as emotion or cognition? A popular explanation of the behavioural effect is that it reflects the modulation of visual cortex by the amygdala, and given its central role in emotion, the effect could be deemed emotional in nature. However, at least five additional brain

mechanisms are involved in prioritising the processing of negative items (Pessoa, 2013). Furthermore, several of them are believed to be central to attention more generally, not just "emotion;" most notably, regions in frontal and parietal cortex that modulate visual processing according to an item's behavioural significance. Indeed, the results of the attentional blink experiment were consistent with both amygdala *and* frontal cortex influencing visual cortical responses (Lim et al., 2009). So, should one consider the role of the frontal cortex cognitive, and that of the amygdala emotional? This approach is unlikely to be productive, however. Both the frontal and the parietal cortex contain a so-called *priority map*, namely, a representation of spatial locations that are salient (such as high-contrast stimuli) or behaviourally relevant (such as stimuli connected to current goals). It is now recognised that in addition to stimulus- and goal-related factors, emotional and motivational variables also determine a stimulus' priority (Pessoa, 2017a). To incorporate biological significance into priority maps, fronto-parietal regions appear to interact with multiple "evaluative" brain regions, including the amygdala, orbitofrontal cortex, and striatum, to prioritise processing based on the emotional/motivational value of a stimulus. Taken together, I argue that to understand even a fairly simple phenomenon as the affective attentional blink, we need to consider the interactions between multiple systems.

I have proposed that the explanatory lifting that is most greatly needed is at the level of *interactions*. Strictly speaking, however, the concept of an interaction is not opposed to the notion of separable entities, as interactions can also refer to distinct variables, processes, or systems that, themselves, produce effects in a non-additive manner. Accordingly, a better term is perhaps *integration*, which connotes sufficient intertwining between the putatively separate systems that their individuation is a linguistic short-cut. To further clarify this point, let's consider the mechanisms of *fear extinction*, where a conditioned stimulus (CS) no longer predicts the unconditioned one, thereby triggering learning of this new relationship, leading to the "extinction" of the conditioned response once elicited by encountering the CS (Figure 2A).

The medial prefrontal cortex (PFC) plays an important role in regulating the amygdala during fear extinction (Morgan, Romanski, & LeDoux, 1993). At first glance, fear extinction appears to fit the scheme of separate entities interacting to generate a new behaviour: cognition (tied to the medial PFC) controlling emotion (tied to the amygdala) in a top-down fashion. Yet, considering the PFC as "top" and the amygdala as "down" does not take into account the richness of the existing neuronal interactions. Multiple cell groups in the amygdala actually project to the medial PFC whose outputs in turn influence amygdala signals. Some studies even have suggested a "top" role for amygdala neurons in the extinction process. The medial PFC is also the target from the hippocampus, an input that potentiates medial PFC signals during extinction. Furthermore, the medial PFC receives substantial inputs from the thalamus, itself a major subcortical-cortical connectivity hub. More generally, during fear extinction - and, in fact, expression - signals from the amygdala, medial PFC, and several additional brain areas (including the orbitofrontal cortex) *collectively* determine whether or not to produce a response (for references regarding this paragraph, see Pessoa, 2017b). I claim that these multi-region interactions afford greater behavioural malleability when responding to threat.

An alternative approach to attempting to explain fear extinction would possibly label each brain region involved in the following manner: amygdala-valence, medial PFC-regulation, hippocampus-context, thalamus-sensory and visceral signals, orbitofrontal cortex-motivation, and so on. One could then describe the ensuing behaviour in terms of the "standard" interactions between the putative processes (valence, regulation, etc.) (Figure 2B). But these processes are not separable; they do not encode stable variables that are simply modulated by other variables. Therefore, explanations in terms of standard interactions will be found wanting - *integration* is needed (Figure 2C). To further develop this point, I will discuss an even broader conceptual issue that informs the understanding of emotion, and in fact of behaviour more generally.

Embracing complexity

Diagrams of mental processes or brain functions, and the general separation of mental domains, are ultimately tied to how scientists understand *causation* in the mind-brain. Stated in simple terms, scientists tend to subscribe to what can be called a *billiard-ball* causal model (Pessoa, 2017a). In this scheme, force applied to a ball leads to its movement on a table until it hits a target ball. The reason the target ball moves is obvious; the first ball hits it, and via the force applied to it, the target ball moves. But this mode of thinking, which has been very productive in the history of science, is too impoverished when complex systems - including the mind and brain - are considered. "Complex systems" are comprised of multiple parts such that interactions between components lead to system-level - so-called emergent - properties that cannot be straightforwardly isolated or attributed to more elementary parts (Mitchell & Newman, 2002). Emotion and motivation are *not* isol-able from perception and cognition. This is *not* to say that mental processes are so interrelated as to become one and the same thing. But when systems are not isolable, interactions and integration of signals are the central elements to be unravelled.

The simplicity of the billiard-ball model depends on the existence of two spatially independent items (billiard balls) that make simple contact with each other. Typical diagrams place mental processes such as motivation and attention in separate boxes (like billiard balls) that can affect each other in direct, simple ways (like a ball hitting another). But this analogy will not be helpful in non-decomposable systems like the mind and brain - and I would argue hinders progress (Figure 2). Thinking of causation in complex systems is much more challenging, of course. Science has only been doing it more seriously since the 1940–1950s (Von Bertalanffy, 1956). For one, we need new ways of creating diagrams, not to mention a new vocabulary and/or way of describing mental and brain phenomena. That, of course, is a very tall order, even if we all agreed that these changes were needed (or, more modestly, useful).

At the most general level, the present discussion speaks to how we should study systems as complex as minds and brains. Dissecting phenomena in terms of their component parts seems like an unimpeachable methodology, to the extent that it can be viewed as almost an axiom of modern science (Deacon, 2011). The shift described here, which has been advocated by many others (for example, Maturana & Varela, 1987; Thompson, 2007), proposes a focus not on parts but on *processes*, which must be understood not solely in terms

of their putative constituent elements but in terms of interactions/integration, as well as their temporal evolution (Pessoa & McMenamin, 2016). Importantly, when "processes" - or more generally complex systems - are invoked there need not be anything vague about it. The science of complexity has progressed enormously in the past 80 years, and continues to evolve rapidly. Thus, the approach advocated here is still "mechanistic" but the focus is inherently interactionist and dynamic.

The study of mind and brain will thus benefit from mathematical and computational tools from complex systems, dynamical systems, and control theory (Csete & Doyle, 2002; Grossberg, 2017; Hirsch, Smale, & Devaney, 2013; Sugihara et al., 2012). Such models will be increasingly important in refining our intuitions about brain and behaviour, not to mention formulating experimental predictions. Several other domains of knowledge also require such "complex systems" approach for progress to be made, including the study of evolution and ecology. The latter is a particularly active area of research, for example, in the study of indirect effects in ecological networks (Guimarães, Pires, Jordano, Bascompte, & Thompson, 2017), higher-order interactions in large ecological communities (Grilli, Barabás, Michalska-Smith, & Allesina, 2017), and causality in complex ecosystems (Sugihara et al., 2012). Importantly, the problems addressed are not simply abstract, or theorerical, ones. For instance, a central objective in ecology is to explain how the tremendous diversity of species in nature persists despite differences in their ability to compete for survival. In a classic paper, the theoretical biologist Robert May (1972) showed formally that, under some assumptions, community diversity destabilises ecological systems. In other words, diverse communities lead to instabilities such as the local elimination of certain species. Recent theoretical results show, however, that higher-order interactions can cause communities with greater diversity to be, instead, more stable than their species-poor counterparts, contrary to classic theory (Bairey, Kelsic, & Kishony, 2016; Levine, Bascompte, Adler, & Allesina, 2017). The goal of this brief excursion into ecology is to illustrate that concrete steps are being taken in other research areas, and that the implications can be indeed foundational.

What type of science of brain and behavior?

The argument outlined here is that researchers should strive to contribute toward the development of a science that explains complex behaviours, not just emotion. Because "interesting behaviors" do not respect the boundaries of research areas, the field should move past defining emotion, or worrying about its rightful place in the mind and brain. I also suggest that serious progress in our "new science" will require embracing both integration and complexity in ways that far outstrip current attempts in this direction. A qualitative jump is required here, one which will necessitate the training of a new generation of researchers equally versed in experimental and formal methods.

Until now, I have side-stepped the question of what constitutes "interesting" or "complex" behaviours, which I have purposively left unaddressed. Adolphs and Andler (2018) suggest that the level of "semi-flexibility" is an appropriate target for emotion research. Although having a target is laudable and can help direct investigation efforts, I am afraid that any one given level is too constraining, and more likely to be a straight-jacket in the end. Ethology -

the science of behaviour, especially in its natural environment - has made vigorous progress in the past few decades. Natural behaviours are nuanced, flexible, context dependent, and plastic, especially so for mammals, not to mention primates and humans. It is time for emotion research to embrace that complexity and let go of more than a century of viewing emotion as tied to primitive, inflexible behaviours. Failure to change course will only enhance the "tunnel vision" that has dominated, for instance, the study of fear (Paré & Quirk, 2017), which has been described as having hit a conceptual cul-de-sac (Kim & Jung, 2018). What has been stated in the case of the study of fear could well pertain to emotion research more generally - unless we detour into more profitable scientific avenues.

Acknowledgements

References

Adolphs R, & Andler D (2018). Investigating emotions as functional states distinct from feelings. Emotion Review, Epub ahead of print.

Anderson BA (2017). Controlled information processing, auto-maticity, and the burden of proof. Psychonomic Bulletin & Review, 1–10. Epub ahead of print. [PubMed: 27368622]

Awh E, Belopolsky AV, & Theeuwes J (2012).Top-down versus bottom-up attentional control: A failed theoretical dichotomy. Trends in Cognitive Sciences, 16(8), 437–443. [PubMed: 22795563]

Bairey E, Kelsic ED, & Kishony R (2016). High-order species interactions shape ecosystem diversity. Nature Communications, 7, 12285.

Clayton NS, & Emery NJ (2015). Avian models for human cognitive neuroscience: A proposal. Neuron, 86(6), 1330–1342. [PubMed: 26087161]

Csete ME, & Doyle JC (2002). Reverse engineering of biological complexity. Science, 295(5560), 1664–1669. [PubMed: 11872830]

Deacon TW (2011). Incomplete nature: How mind emerged from matter. New York, NY: W. W. Norton & Company.

Fox AS, Lapate RC, Davidson RJ, & Shackman AJ (Eds.). (2018). The nature of emotion: Fundamental questions. New York, NY: Oxford University Press.

Grilli J, Barabás G, Michalska-Smith MJ, & Allesina S (2017). Higher-order interactions stabilize dynamics in competitive network models. Nature, 548(7666), 210–213. [PubMed: 28746307]

Grossberg S (2017). Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. Neural Networks, 87, 38–95. [PubMed: 28088645]

Guimarães PR, Jr, Pires MM, Jordano P, Bascompte J, & Thompson JN (2017). Indirect effects drive coevolution in mutualistic networks. Nature, 550(7677), 511–514. [PubMed: 29045396]

Hirsch MW, Smale S, & Devaney RL (2013). Differential equations, dynamical systems, and an introduction to chaos (3rd ed.). Waltham, MA: Academic press.

Kim JJ, & Jung MW (2018). Fear paradigms: The times they are a-changin'. Current Opinion in Behavioral Sciences, 24, 38–43. [PubMed: 30140717]

Levine JM, Bascompte J, Adler PB, & Allesina S (2017). Beyond pairwise mechanisms of species coexistence in complex communities. Nature, 546(7656), 56–64. [PubMed: 28569813]

Lim S-L, Padmala S, & Pessoa L (2009). Segregating the significant from the mundane on a moment-to-moment basis via direct and indirect amygdala contributions. Proceedings of the National Academy of Sciences of the United States of America, 106(39), 16841–16846. [PubMed: 19805383]

Maturana HR, & Varela FJ (1987). The tree of knowledge: The biological roots of human understanding. Boston, MA: New Science Library/Shambhala Publications.

May RM (1972). Will a large complex system be stable? Nature, 238(5364), 413–414. [PubMed: 4559589]

Melnikoff DE, & Bargh JA (2018). The mythical number two. Trends in Cognitive Sciences, 22(4), 280–293. [PubMed: 29571664]

Mitchell M, & Newman M (2002). Complex systems theory and evolution In Pagel M (Ed.), Encyclopedia of evolution (pp. 1–5). New York, NY: Oxford University Press.

Moors A, Boddez Y, & De Houwer J (2017). The power of goal-directed processes in the causation of emotional and other actions. Emotion Review, 9(4), 310–318.

Morgan MA, Romanski LM, & LeDoux JE (1993). Extinction of emotional learning: Contribution of medial prefrontal cortex. Neuroscience Letters, 163(1), 109–113. [PubMed: 8295722]

Paré D, & Quirk GJ (2017). When scientific paradigms lead to tunnel vision: Lessons from the study of fear. Npj Science of Learning, 2(1), 6. [PubMed: 30294453]

Pessoa L (2005). To what extent are emotional visual stimuli processed without attention and awareness? Current Opinion in Neurobiology, 15, 188–196. [PubMed: 15831401]

Pessoa L (2013). The cognitive-emotional brain: From interactions to integration. Cambridge, MA: MIT press.

Pessoa L (2017a). Cognitive-motivational interactions: Beyond boxes-and-arrows models of the mind-brain. Motivation Science, 3(3), 287–303. [PubMed: 29399604]

Pessoa L (2017b). A network model of the emotional brain. Trends in Cognitive Sciences, 21(5), 357–371. [PubMed: 28363681]

Pessoa L, & McMenamin B (2016). Dynamic networks in the emotional brain. The Neuroscientist, Epub ahead of print.

Rischawy I, Blum M, & Schuster S (2015). Competition drives sophisticated hunting skills of archerfish in the wild. Current Biology, 25(14), R595–R597. [PubMed: 26196481]

Shallice T (1988). From neuropsychology to mental structure. Cambridge: Cambridge University Press.

Sugihara G, May R, Ye H, Hsieh CH, Deyle E, Fogarty M, & Munch S (2012). Detecting causality in complex ecosystems. Science, 338(6106), 496–500. [PubMed: 22997134]

Thompson E (2007). Mind in life: Biology, phenomenology, and the sciences of the mind. Cambridge, MA: Harvard University Press.

Von Bertalanffy L (1956). General system theory. General Systems, 1,1–10.

Vuilleumier P (2005). How brains beware: Neural mechanisms of emotional attention. Trends in Cognitive Sciences, 9(12), 585–594. [PubMed: 16289871]
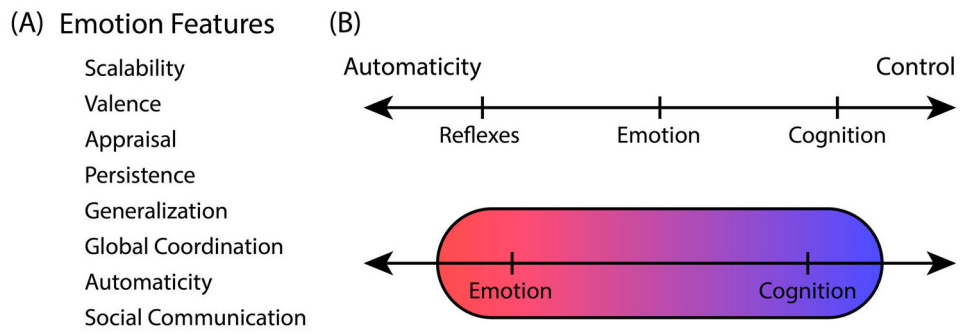
Figure 1.
Emotion properties. (A) Suggested key properties of emotion by Adolphs and Andler (2018). (B) The axis of automaticity from stronglyautomatic to highly controlled. Top: Emotional phenomena are purported to be situated between reflexes and cognitive processes. Bottom: An alternative view in which emotional and cognitive phenomena largely overlap along the automaticity axis.
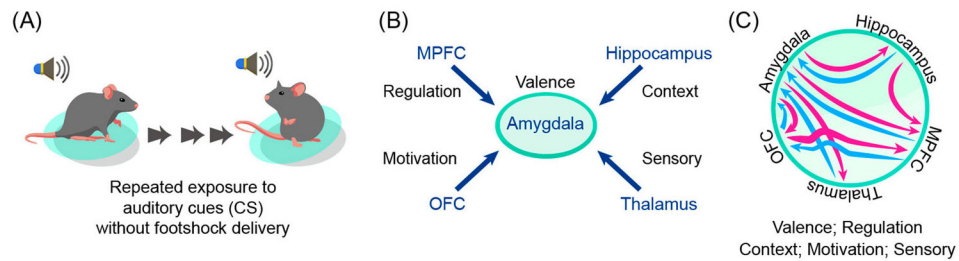
Figure 2.
Fear extinction and brain processes. (A) Schematic representation of fear extinction. (B) Conceptualisation of fear extinction in terms of the top-down regulation of the amygdala by the medial prefrontal cortex, with additional variables influencing the process. (C) Schematic representation of the connections between some of the brain regions involved, emphasising a non-hierarchical view of the processes leading to fear extinction. The descriptors "valence," "regulation," and so on, are not tied to brain areas in any straightforward one-to-one fashion. Abbreviations: CS, conditioned stimulus; MPFC, medial prefrontal cortex OFC, orbitofrontal cortex.