ARTICLE

emotion review

Emotion Review
1–8

ISSN 1754-0739
DOI: 10.1177/1754073919843185
https://journals.sagepub.com/home/emr

# Rethinking the Principles of Emotion Taxonomy

Assaf Kron
*Department of Psychology, University of Haifa, Israel*

## Abstract

This article examines whether a functionalist approach to emotion classification is a research program that can feasibly be implemented in an experimental environment. I suggest that this is a promise perhaps impossible to keep. The crux of the argument is that if functional taxonomy is to go the full distance and shape experimental conditions to the new boundaries, then stimuli/experimental manipulations must be selected based on functional principles. But this seems implausible or even impossible. I conclude that emotion taxonomy, and thus the metastructure of emotion theory, is both constrained and determined by human emotional feelings, and suggest that feelings have a special epistemological status in emotion research.

**Keywords**
classification, emotion taxonomy, functionalism

## The Argument in a Nutshell

Recent theoretical approaches have emerged that are calling for emotions to be classified according to their function (Adolphs & Andler, 2018; LeDoux, 2012, 2013, 2014, 2017). This is in response to what was perceived as the unjustified prominent role that feelings (conscious experience of emotions) play in constructing emotion taxonomy. The extreme version of this reaction is the suggestion that the science of emotion should, at least temporarily, omit feelings (Adolphs & Andler, 2018). In the present article, I examine if a functionalist approach to emotion classification is a research program that can feasibly be implemented in an experimental environment. I suggest that this is a promise perhaps impossible to keep. Critically, experimental research should be able to manipulate emotions, and to do so, a stimulus or manipulation must be selected. I argue that it is extremely challenging to select stimuli or experimental manipulations based purely on functional principles (in contrast to feelings-based principles). The crux of the argument is that, in order to escape the vicious circle of feelings-based taxonomy (and replace it with functional principles), stimuli/experimental manipulations must be selected based on functional principles. But this seems implausible or even impossible. I conclude that emotion taxonomy, and thus the metastructure of emotion theory, is both constrained and determined by human emotional feelings, and suggest that feelings have a special epistemological status in emotion research.

## Introduction

Traditionally, affective science has classified emotions using a feelings-based taxonomy—that is, the categories scientists use when classifying emotions are the same categories used to classify emotional feelings (e.g., Adolphs & Andler, 2018; LeDoux, 2017). Categories such as pleasure, displeasure, fear, anger, sadness, and happiness are used both to classify emotions and affect in scientific research as well as emotional feelings. Feelings-based taxonomy has several drawbacks, which has been discussed elsewhere (e.g., Adolphs & Andler, 2018; LeDoux, 2017). Given these limitations, recent theoretical moves suggest that, in order to improve scientific investigation, it would be better to classify emotions in terms of their hypothesized evolutionary function instead of conscious feelings (Lang, 1994, 2010; Lang, Bradley, & Cuthbert, 1998; LeDoux, 2012, 2013, 2014, 2017; see also Miskovic, Kuntzelman, & Fletcher, 2015, for related review). This is inspired by functionalism, a theory that defines mental states based on their

*Author note*: I would like to thank Oksana Itkes and Dr. Baruch Eitam for their questions during a lecture, which I won't forget. They left me with no choice but to write this article. I would also like to thank them for their helpful comments on the manuscript. I would like to thank, Prof. Amit Bernstein, Prof. Nachshon Meiran, and Ms Orit Heimer for their helpful comments on the manuscript, and Ms Leigh Hall for proofreading and editing it.

*Corresponding author*: Assaf Kron, Department of Psychology, The Institute of Information Processing and Decision Making, University of Haifa, aba Hushi, Haifa 31905, Israel.
*Email*: akron@psy.haifa.ac.il

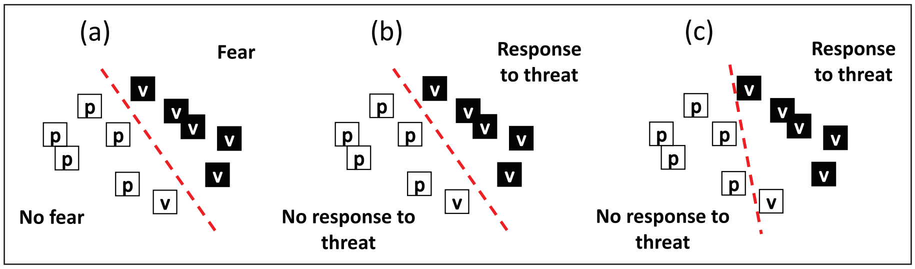


**Figure 1.** Figures 1a to 1c depict three instances of the same phenomenon space. The squares in each space represent phenomena. Each phenomenon has two features: some are white and some are black; some are "v" and some are "p." The researcher uses the broken red line to divide the phenomenological space into two groups. Figure 1a represents the condition before a terminological or epistemic change. Figure 1b depicts a terminological change. Figure 1c depicts terminological and epistemic changes.

functional role (Polger, 2012). In the case of emotions, it suggests that omitting feelings-based taxonomy and replacing it with a set of categories that reflect the function of emotions would be beneficial. For example, instead of using the term *fear*, which reflects feelings-based categorization, it suggests using the phrase *detection of* or *response to threat*, which reflects the assumed evolutionary function of fear. In the following argument, I examine the feasibility of the functional approaches to emotion classification. I begin by explaining what emotion taxonomy is and its importance. In what follows, in order to be able to discuss the potential functional change in the taxonomy, a distinction is made between two types of changes: *terminological change* (changing the titles of categories) and *epistemic change* (changing the boundaries between the categories).

## Emotion Taxonomy

Emotion taxonomy is an emotion classification system. Scientists use emotion classification systems, sets of categories, to classify emotions. Examples of emotion taxonomies include the "basic emotions" classification system (Ekman, 1992) that suggests that the affective space is best divided into six discrete categories: anger, disgust, fear, happiness, sadness, and surprise; and Cowen and Keltner (2017), who suggest that affective space be divided into 27 discrete categories including sadness, horror, and awkwardness. Other types of taxonomies are dimensional, that is, they do not classify emotions into discrete categories, but according to their relative locations in respect to continuous dimensions. For example, Schlosberg (1954) suggested three dimensions of emotion: pleasantness–unpleasantness, attention–rejection, and level of activation.

For better or for worse, the emotion classification system shapes the metastructure of emotion theory and experimentation. On the one hand, it provides scientists with the language to ask questions about emotions and, on the other hand, it construes and constrains the affective space into a specific categorical structure, thus preventing scientists from asking questions about categories that are not included in that classification system. It is not surprising that the debates about emotion classification are long and intense (e.g., Barrett, 2006; Russell, 1980), as they are not only about emotion classification itself, but also the nature of emotion research programs and the structure of emotion theory in general.

## Terminological Change Versus Epistemic Change

The functional approach to emotion classification (e.g., Adolphs & Andler, 2018) calls to change the emotion classification system in such a way that the categorical structure will reflect different functions and not different feelings. In order to make evaluating the proposed functional change possible, I first specify two types of potential changes that could be made to the classification system: terminological changes, that is, changing the title of categories, and epistemic changes, defined here as changing the boundaries between categories. Figure 1 illustrates the difference between terminological and epistemic changes. Figures 1a to 1c represent the same space. This space contains 11 squares, each representing a different phenomenon. The researcher uses a classifier (the broken red line) to divide the space into two groups. Figure 1a represents the condition before a terminological or epistemic change, and the space is demarcated with black versus white boxes. In Figure 1b, the researcher also uses a classifier that demarcates the space with black versus white boxes. This time, however, one group is titled *response to threat* and the second group is titled *no response to threat.* In this second case, the researcher uses new terminology, but the space is parsed using the same demarcations—black or white boxes. Figure 1c describes both a terminological and an epistemic change, the space is parsed using a different principle, and each group is given the new title. Using the distinction between terminological and epistemic changes, the main question the remainder of this article will attempt to answer is whether or not a functional classification can result in an epistemic change.

## Stimulus Selection Rule: A Criterion for Epistemic Change

To answer the question of whether or not the proposed functional classification can result in an epistemic change, we should examine if it is possible for emotion researchers to use *function* as a criterion to actually classify a phenomenon (and not only as a new set of titles for emotion categories). And to answer this, we should first determine precisely what the phenomena to be classified are. Classification can be used at different levels of abstraction, for example, at the theoretical level (e.g., to determine the emotion categories) or the experimental level (e.g., to classify experimental conditions and/or stimuli or manipulations used to elicit emotions in experiments). I suggest that in order to move from ambition to reality, the functional classification should go "the whole nine yards" and result in an epistemic change in the structure of experimental design.

What does it mean to make an epistemic change in the structure of the experimental design? Taxonomy shapes the experimental design. In many cases, the scale of the independent variable implicitly or explicitly reflects the assumed emotion classification system. For example, experimental conditions that manipulate fear versus disgust assume a classification system that has categories of fear and disgust. The emotion categories that are carved by the experimental conditions shape the data to be collected, constrain the scope of analyses, and construe conclusions and the aggregation of empirical knowledge. This is why it is critical that the functional epistemic change does not remain at the conceptual level, but will go all the way to shape the experimental conditions to the new boundaries. If the new classification system does not shape the boundaries of the experimental conditions, aggregation of data (and hence scientific knowledge) will remain in terms of and reflect the structure of the old classification system.

How could we determine if the new, functional classification system is able to restructure the boundaries between experimental conditions? Allegedly, the most intuitive test is to examine if the labels of the experimental conditions match those of the classification system (functional categories). However, altering the labels could potentially merely reflect a change in terminology, while the actual boundaries between categories remain unchanged. Here I suggest a simple rule to decide if a new classification system (functional in our case) has made it all the way to shaping the boundaries between the experimental conditions: A classification system can reshape the boundaries between experimental conditions if there is a function-based classifier that can be used to select stimuli for experimental conditions.

The logic of this rule is that the criterion used to select stimuli is the most concrete reflection of the classification system. Accordingly, the crux of the decision rule of whether a functional classification system can shape the boundaries between the experimental conditions is the ability of the researcher to select stimuli according to their function, and not according to the feelings they elicit. Consider a researcher who works according to the feelings-based classification system. She is interested in examining the differences between fear and anger and plans to use movie clips as emotion elicitation stimuli. How will the researcher a priori decide which clips will elicit fear and which will elicit anger? One option is to watch the clips herself, reflect on her own emotional feelings, and allocate the clips accordingly. Alternatively, she can run a pilot study asking participants how much fear and/or anger they feel while watching the clips, and then allocate the stimuli according to the average ratings. In both cases, stimuli were allocated to experimental conditions according to feelings—termed here as a "feelings-based stimulus classifier." Now, consider a functionalist researcher who is not interested in the difference between fear and anger, but between a response to threat and a response to deliberate, unfair treatment. This functional definition partially overlaps with fear and anger, but the function-based stimulus classifier will not use feelings, but instead will classify the clips according to whether or not stimuli are perceived as a threat and/or deliberate unfair treatment.

As yet, most stimuli are selected according to feelings-based classifiers; one indication is the standardized stimulus pools such as video clips (e.g., Gross & Levenson, 1995), pictures (e.g., International Affective Picture System [IAPS]; Lang, Bradley, & Cuthbert, 2008), and audio recordings (e.g., Bradley & Lang, 1999), and words (e.g., Bradley & Lang, 2007). Most stimulus pools are organized according to self-reports of feelings (see also Dan-Glauser & Scherer, 2011; Gross & Levenson, 1995; Marchewka, Żurawski, Jednoróg, & Grabowska, 2014). This implies that a feelings-based stimulus classification is being used in studies relying on these norms to choose stimuli.

In summary, the question of whether the new functional classification system can go all the way and reshape the boundaries of the experimental conditions is now more specific: is there a way to select stimuli according to their function (and not according to feelings)? The clear advantage of stimulus selection as a rule (to determine if the classification system can reshape the borders of the experiment conditions) is that it reflects the assumed classification system in the design and can promise that the change is not merely terminological.

## The Functional Classifier

I will now examine the feasibility of the functional classification system—is there a way to allocate stimuli to experimental conditions according to functional principle? As a case study, this argument focuses on the negative valence pole with specific attention to fear response. I focus on fear because it already has well-developed functional alternatives and I believe it is the most plausible option for the functional move. One well-discussed functional alternative to fear is *a response to a threatening* or *noxious event* (LeDoux, 2014). The narrower version of our main question is thus if a feelings-based classifier (e.g., classifying stimuli in experimental conditions based on the feelings they elicit—for example, fear vs. no fear) can be replaced by a functional classifier (e.g., classifying stimuli in experimental conditions as threat vs. no threat). Note that the functional definition is *a response to threat* and therefore stimuli must be classified into groups of threat versus no threat, so when perceived by participants they will elicit a response to threat versus no response to threat. The remainder of this article will thus focus on the narrower question: is there a feelings-free classifier that can categorize stimuli into groups of "threat" versus "no threat"?

As far as I know, there is still no research using a feelings-free stimulus classifier that supports functional classification. For the sake of discussion, I have created a fictitious "functionalist researcher" that is attempting to select stimuli for the experiment conditions. The functionalist researcher will need to select stimuli for experiments according to whether or not they are a threat. The challenge is that stimuli must be selected without any knowledge of whether or not the stimuli elicit feelings of fear. The functionalist cannot define threat as something that people are afraid of, because this would be considered a feelings-based classifier. The logic behind the evaluation process in what follows might seem counterintuitive to the emotions researcher at first, because, in emotions literature, stimuli or manipulations are generally assumed to be a threat because they elicit fear. Instead, the functionalist should be able to accurately classify events as either threatening or nonthreatening according to a specific "feelings-free" feature.

## Evaluation of the Functional Classifier

Here I will examine four stimulus classifiers for human emotions research and three for animal research. Although they are not the only potential classifiers, I chose to focus on them for two reasons: several of them had been suggested previously (e.g., LeDoux, 2014), and because I believe they are the strongest classifiers and thus provide the best prospects for the functional alternative.

### *Functional Classifiers in Human Research*

The four classifiers for human research I will discuss are heuristics, biological markers, approach–avoidance motor behavior, and analytic evaluation.

***Heuristics.*** Perhaps the most straightforward functional classifier is a simple heuristic: intuitively judging each stimulus as either threatening or nonthreatening, as is done in standard superordinate classification tasks. The shortcoming of such a classification is that intuitive judgments may be completely or partially biased due to the researcher's own feelings or knowledge about the expected feelings related to a particular stimulus. This potential bias reflects a feelings-based classifier. There is much evidence suggesting that judgment is influenced by emotional feelings and mood (for reviews, see Schwarz, 2000; Schwarz & Clore, 1988, 1996), and that individuals are more likely to retrieve information that is congruent with their current feelings (e.g., Bower, 1981; Isen, Shalker, Clark, & Karp, 1978).

A second potential problem with an "intuitive functional classifier" based on the researcher's heuristic is that feelings can also sneak into the researcher's judgment unintentionally via an empathic stance, such as simulating participants' feelings. For example, consider a functionalist researcher attempting to decide if an electric shock is a threat to human participants. The researcher's judgment may be based on imagining what it would be like to receive a shock as a participant. The empathic stance, in turn, might involve the simulation of shock and result in the observer feeling pain or thinking about feeling pain (e.g., Jackson, Meltzoff, & Decety, 2005; Jackson, Rainville, & Decety, 2006; Lamm, Decety, & Singer, 2011; Singer et al., 2004).

***The biological marker.*** This criterion was suggested as a replacement for feelings-based classifications with an emphasis on the neural circuits that are responsible for a specific function. For example, instead of using the term fear, it has been suggested to refer directly to the neural circuits that underlie the defense response to threat (LeDoux, 2014). Could biological markers (e.g., neural circuits, peripheral autonomic activation, etc.) be used by the functionalist researcher to classify stimuli as threat versus no threat? Regardless of whether or not this scenario could be carried out regularly as part of a research routine, my aim is purely to discuss its logical feasibility.

Selecting stimuli according to biological markers is doomed to go around in a vicious circle. Before selecting threatening stimuli according to a specific neural activation, you should first know which neural activation is associated with the response to threat. This would require the researcher to choose stimuli that are perceived as a threat, test hypotheses, and conclude which neural activations are associated with threat responses; all independent of neural activation. The same logic applies to all biological markers such as hormones, peripheral autonomic changes, and skeletal-motor responses. They cannot all serve as experimental classifiers for threat without getting caught in a vicious circle. In that sense, identifying neural circuits, or any other biological markers associated with threat, cannot solve epistemic dependence on feelings.

***Approach–avoidance motor behavior.*** Can approach–avoidance behavior serve as a stimulus classifier that is feelings-free and also supports a functional classification system? Approach–avoidance instigation can be treated as isomorphic to motor behavior (Schneirla, 1959); for example, John moved his hand toward the chocolate cake and jerked it back when he noticed a cockroach on the plate. Besides motor behavior, approach–avoidance can also be considered a psychological instigation (Elliot & Covington, 2001; Kron et al., 2014); for example, avoiding a meeting by not attending it (i.e., avoiding by not approaching), or approaching a friend during a meeting by not leaving and extending the meeting (i.e., approaching by not avoiding; Elliot, 2006). The logic of using an approach–avoidance classifier comes from the close theoretical relationship between emotion and action; that is, survival depends on actions in response to threat or reward (Lang, 2010): hugging, embracing, clasping, caressing, kissing, hitting, fleeing, screaming, spitting, and baring teeth all involve actions. Emotion, in this context, can be thought of as organized around actions or action tendencies (Frijda, 1988)—that is, motoric action or only the representation of action. Some theories see action as one component of the emotional response (e.g., Russell, 2003), or see emotions as the cause of action (Blakemore & Vuilleumier, 2017; Nanay, 2017). As such, classifiers that are based on actions and action tendencies are potential candidates for replacing feelings-based classifiers.

While many models suggest that behavior is organized around approach or avoidance vectors—that is, that humans approach rewards and avoid threats (see Cacioppo & Berntson, 1994; Corr, 2013; Elliot, Eder, & Harmon-Jones, 2013, for reviews)—empirical findings suggest a much more complex link between threats/rewards and approach/avoidance behavior. People do not always avoid the threat; in fact, they sometimes approach it (Bradley, Cuthbert, & Lang, 1991; Cuthbert, Bradley, & Lang, 1996; Hamm, Cuthbert, Globisch, & Vaitl, 1997; Kron, Pilkiw, Banaei, Goldstein, & Anderson, 2015; Lang, Greenwald, Bradley, & Hamm, 1993; Patrick, Bradley, & Lang, 1993; Vrana, Spence, & Lang, 1988), for example, during a defensive fight or offensive attack, a human or animal might actually approach a threat (Carver & Harmon-Jones, 2009). In addition, sensitivity to reward and punishment is at least partially mediated by personality traits and contextual factors (Corr, 2013; Corr, DeYoung, & McNaughton, 2013, for review). Thus, although the link between avoidance behavior and threat is frequently assumed, there is no empirical evidence to suggest a simple direct link. Using approach–avoidance behavior to determine if a stimulus is a threat is, therefore, less than optimal.

***Probabilistic definition.*** To avoid problems associated with heuristics, biological, and behavioral markers, I will now examine the possibility of defining threat analytically and then using that definition to classify stimuli. I used the following probabilistic working definition of threat: X is a threat to Y if X is likely to cause damage, injury, or harm to Y. According to this working definition, the functionalist researcher should simply decide if stimuli are likely to cause harm or not.

The definition of threat is probabilistic in the sense that it refers to the probability that X will cause damage in the future. Probability can be computed objectively (e.g., as the limit of relative frequency), but it can also be estimated subjectively, namely, intuitively. In the following section, I suggest that it is very challenging and perhaps impossible to use objective probability to select stimuli. That leaves us with subjective probability, which brings us back to the limitations of heuristics as discussed before.

Using objective probability as a classification rule for stimuli is allegedly a promising way that can be used as a benchmark for functional classification. However, looking into the details of using objective probability will reveal a complex, perhaps undoable stimulus selection process that can lead to bizarre classifications and, critically, the inability to validate the stimuli that were chosen. For example, in terms of probability, cars are more likely to cause damage to humans than cockroaches (many more people get hurt by cars than by cockroaches). Imagine a functionalist researcher (Researcher 1) trying to select pictorial emotional stimuli for an experiment. If the functionalist researcher relies solely on objective probabilities, she might find herself concluding that a picture of a new sports car should be allocated to the group of threatening stimuli, and a picture of a cockroach to the group of less threatening stimuli. This decision, from the perspective of feelings-based classification, might seem strange—pictures of cockroaches usually elicit more negative feelings than pictures of sports cars that, in general, elicit positive feelings (Lang, Bradley, & Cuthbert, 1997). Critically, there is no way that we can validate the decision; if we decide to validate manipulations according to self-reports, we are back to feelings-based classification.

The lack of manipulation validation can quickly result in a Tower of Babel of functional arguments with no way to decide between them. For example, imagine Researcher 2 sees Researcher 1's design and completely disagrees with its logic: "a parked sports car is not a good stimulus—the sports car should appear to be approaching the participant in order to be perceived as a threat." Researcher 3 would say "both Researcher 1 and 2 are wrong, only real cars (and cockroaches) approaching the participants could serve as a threat." Researcher 4 will argue "cockroaches are evolutionarily associated with diseases and plague and hence they will be perceived as a greater threat than cars." Unfortunately, we cannot use feelings to decide who is correct. Even though the arguments of Researchers 1, 2, 3, and 4 are all feelings-free, there is no clear criterion to decide between them.

The example of the car and the cockroach illustrates that in order to decide between the functional arguments, we should examine if stimuli were *perceived* as a threat. How will the functionalist researcher decide if stimuli are perceived as a threat without checking the feelings they elicit? When I discuss this with my colleagues, they usually suggest using biological or behavioral markers instead of feelings. But remember that vicious circle? In a "functional world," in order to decide between two functional arguments by a marker (e.g., behavioral, biological), the functionalist needs to know which markers indicate threat, and in order to know which markers indicate threat, the functionalist needs to select the right stimulus.

## *Functional Classification in Animal Research*

So far, this article has focused on the feasibility of adopting a functional classifier in human research. But what about adopting a functional model in animal research, where it was originally developed (Miskovic et al., 2015)? The logic behind testing the functional approach in animal research is identical to that of human research: is there a feelings-free classifier that a researcher can use to select stimuli and manipulations for experimental conditions? My position here is similar to that of human research, with an additional drawback that we also use human feelings in the feelings-based classifiers to classify stimuli or manipulations in animal research. I will discuss three classifiers: anthropomorphic behavior, species-typical responses to threat, and analytical evaluation.

***Anthropomorphic behavior.*** A group of potential classifiers can use knowledge about human behavior and feelings as a basis for classifying stimuli and manipulations in animal research into threat versus no threat categories. Two such classifiers are human behavior and human feelings. In the first case, a functionalist researcher could classify a manipulation (e.g., electric shock) as a threat to rats, since rats demonstrate behavior similar to that of humans in response to threat—for example, both rats and humans scream in response to an electric shock. In the second, the researcher could use human feelings to classify stimuli, speculating what the human's response will be to the electric shock. The main problem with anthropomorphic classifiers is that they inherit, by definition, the problems of the human feelings-based classifiers, with the additional disadvantage of using human feelings-based classifiers to model animal emotions.

***Species-typical responses to threat.*** The second group of classifiers that can potentially be used to classify stimuli and manipulations are species-typical response categories. For example, an electric shock or a stimulus that is associated with an electric shock may cause rats to withdraw (Karpicke, Christoph, Peterson, & Hearst, 1977), freeze (Bouton & Bolles, 1980), or crouch (Blanchard & Blanchard, 1969). If such rat-typical responses can serve as a classifier, the researcher should be able to use it to determine if a stimulus or manipulation is rewarding or threatening (it can be tempting to believe that we can identify species-typical responses to threat regardless of our feelings; LeDoux, 2014). However, this group of classifiers falls into the same vicious circle as described before for biological markers; in order to use species-typical responses to a threat as a classifier for experiment manipulation, the functionalist researcher should first know which behaviors are species-typical responses to threat. Threat must first be manipulated in the experiment to a priori select manipulations that elicit threat versus no threat.

***Probabilistic definition.*** Is it possible to define threat analytically (X is a threat if X is likely to cause damage, injure, or harm) and then use that definition to classify stimuli? This potential classifier has the same problems as in human research plus an additional level of complexity. In order to respond to threat, the animal must *perceive* the object or event as a threat or that it will cause damage. Thus, like in humans, the question is not only which test the researcher should use to decide if a stimulus could be physically harmful to the animal, but also how the researcher can know that the animal *perceives* a stimulus as threatening. The functionalist researcher needs a marker to decide if the animal *perceived* the manipulation or stimulus as a threat. Using anthropomorphized behavior, species-typical behavior, or biological markers brings the functionalist back to the all-too-familiar vicious circle.

## Discussion

This article examines the feasibility of a functional classification system for human and animal emotions research. More specifically, it ponders whether functional classification can result in epistemic change.

The first part of the article was devoted to formulating a benchmark to decide if epistemological change was achieved. While a terminological change is easy to detect—the terms are different—epistemic change can be more elusive. In the context of empirical research, epistemological change can be achieved only when there is the possibility to choose stimuli or manipulations accordingly. Epistemic change is defined at the level of stimuli selection (instead of at a more abstract level) because experimental design construes and limits the entire structure of data collection, analysis, and conclusion. If change cannot be implemented at the level of experimental design, it will not affect the structure by which knowledge is aggregated.

With regard to an epistemic change, the position presented in this article is that a functional classifier is very challenging and perhaps impossible. When it comes to details regarding how the classifier would be implemented to select stimuli using functional features, the argument I presented suggests that the main alternatives offered in the literature all require an a priori definition of threat that relies on feelings, or in any case, are not completely feelings-free.

One limitation of this article is its focus on one single potential functional move—*fear* versus *response to threat*—while others (e.g., *sadness* vs. *response to loss*) were excluded. There are two reasons for this. First, the response to threat is believed by many to be a fundamental "defense circuit" that can be traced in the phylogenetic tree. Second, there were already previous suggestions to replace fear with a functional classification. Nevertheless, focusing on just one type of response, the argument articulates two principles that could be used in the future to test the feasibility of other functional moves or other (nonfunctional) models that claim to be independent of conscious feelings. The first principle distinguishes between epistemic and terminological changes: epistemic change is only possible if a model has an appropriate classifier with which to allocate stimuli/manipulations to experimental conditions. The second principle is that classifiers should be "feelings-free" and should not be dependent on the researcher's judgment.

A second limitation of this article is the scope of classifiers that were examined. The classifiers reflect the main implicit and explicit alternative models found in the literature, including heuristics, biological markers, approach–avoidance behavior, and probabilistic definition of threat. While it clearly does not cover all possible classifiers, the logic in this argument could be used to test potential classifiers in the future.

Finally, the current argument accepts the limitations associated with feelings-based classifiers (Adolphs & Andler, 2018), and at the same time it also hampers the enthusiasm for a functional classification system, arguing that it does not result in epistemic change. You, the reader, may feel you have been left empty-handed as the argument offers no solution. But despite its rather pessimistic tone, it also offers some positive and constructive messages. It illuminates the critical role that models of human emotional experiences have in emotion classification systems, it portrays how classification systems mediate and construe data collection and aggregation of conclusions in both human and animal emotion research, and it draws attention to the assumptions that are needed in order to use feelings-based taxonomy in modeling unconscious phenomena.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### References

Adolphs, R., & Andler, D. (2018). Investigating emotions as functional states distinct from feelings. *Emotion Review*, *10*, 191–201.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46.

Blakemore, R. L., & Vuilleumier, P. (2017). An emotional call to action: Integrating affective neuroscience in models of motor control. *Emotion Review*, *9*, 299–309.

Blanchard, R. J., & Blanchard, D. C. (1969). Crouching as an index of fear. *Journal of Comparative and Physiological Psychology*, *67*(3), 370–375.

Bouton, M. E., & Bolles, R. C. (1980). Conditioned fear assessed by freezing and by the suppression of three different baselines. *Animal Learning & Behavior*, *8*(3), 429–434.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148.

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1991). Startle and emotion: Lateral acoustic probes and the bilateral blink. *Psychophysiology*, *28*(3), 285–295.

Bradley, M., & Lang, P. J. (1999). *The International Affective Digitized Sounds (IADS) stimuli, instruction manual and affective ratings*. NIMH Center for the Study of Emotion and Attention.

Bradley, M. M., & Lang, P. J. (2007). *Affective Norms for English Text (ANET): Affective ratings of text and instruction manual* (Technical Report D-1). Gainesville, FL: University of Florida.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401–423.

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin*, *135*(2), 183–204.

Corr, P. J. (2013). Approach and avoidance behaviour: Multiple systems and their interactions. *Emotion Review*, *5*, 285–290.

Corr, P. J., DeYoung, C. G., & McNaughton, N. (2013). Motivation and personality: A neuropsychological perspective. *Social and Personality Psychology Compass*, *7*(3), 158–175.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology*, *33*(2), 103–111.

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva Affective Picture Database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*(2), 468–477.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Elliot, A. J. (2006). The hierarchical model of approach–avoidance motivation. *Motivation and Emotion*, *30*(2), 111–116.

Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. *Educational Psychology Review*, *13*(2), 73–92.

Elliot, A. J., Eder, A. B., & Harmon-Jones, E. (2013). Approach–avoidance motivation and emotion: Convergence and divergence. *Emotion Review*, *5*, 308–311.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*(5), 349–358.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, *9*(1), 87–108.

Hamm, A. O., Cuthbert, B. N., Globisch, J., & Vaitl, D. (1997). Fear and the startle reflex: Blink modulation and autonomic response patterns in animal and mutilation fearful subjects. *Psychophysiology*, *34*(1), 97–107.

Isen, A. M., Shalker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, *36*(1), 1–12.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, *24*(3), 771–779.

Jackson, P. L., Rainville, P., & Decety, J. (2006). To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain*, *125*(1–2), 5–9.

Karpicke, J., Christoph, G., Peterson, G., & Hearst, E. (1977). Signal location and positive versus negative conditioned suppression in the rat. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*(2), 105–118.

Kron, A., Pilkiw, M., Banaei, J., Goldstein, A., & Anderson, A. K. (2015). Are valence and arousal separable in emotional experience? *Emotion*, *15*(1), 35–44.

Kron, A., Pilkiw, M., Goldstein, A., Lee, D. H., Gardhouse, K., & Anderson, A. K. (2014). Spending one's time: The hedonic principle in ad libitum viewing of pictures. *Emotion*, *14*(6), 1087–1101.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, *54*(3), 2492–2502.

Lang, P. J. (1994). The varieties of emotional experience: A meditation on James–Lange theory. *Psychological Review*, *101*(2), 211–221.

Lang, P. J. (2010). Emotion and motivation: Toward consensus definitions and a common research purpose. *Emotion Review*, *2*, 229–233.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Retrieved from https://www2.unifesp.br/dpsicobio/adap/instructions.pdf

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*(12), 1248–1263.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8. (Gainesville, FL: University of Florida.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273.

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, *73*, 653–676.

LeDoux, J. E. (2013). The slippery slope of fear. *Trends in Cognitive Sciences*, *17*(4), 155–156.

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences*, *111*(8), 2871–2878.

LeDoux, J. E. (2017). Semantics, surplus meaning, and the science of fear. *Trends in Cognitive Sciences*, *21*(5), 303–306.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610.

Miskovic, V., Kuntzelman, K., & Fletcher, N. (2015). Searching for affect in affective neuroscience: Challenges and opportunities. *Psychology of Consciousness: Theory, Research, and Practice*, *2*(1), 1–18.

Nanay, B. (2017). All actions are emotional actions. *Emotion Review*, *9*, 350–352.

Patrick, C. J., Bradley, M. M., & Lang, P. J. (1993). Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology*, *102*(1), 82–92.

Polger, T. W. (2012). Functionalism as a philosophical theory of the cognitive sciences. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 337–348.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review*, *61*(2), 81–88.

Schneirla, T. C. (1959). An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation 1959* (pp. 1–42). Lincoln: University of Nebraska Press.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*(4), 433–440.

Schwarz, N., & Clore, G. L. (1988). How do I feel about it? The information function of affective states. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior: New evidence and integrative attempts* (pp. 44–62). Toronto, Canada: Hogrefe & Huber.

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 385–407). New York, NY: The Guilford Press.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*(5661), 1157–1162.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: A new measure of emotion? *Journal of Abnormal Psychology*, *97*(4), 487–491.