To cite this chapter:

Moors, A. (in press). What is an emotion? In O. J. Braddick (Ed.), *Oxford Research Encyclopedias: Psychology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.918

What is an Emotion? Agnes Moors KU Leuven – forthcoming in Oxford Research Encyclopedia of Psychology

Summary

To answer the question "What is an emotion?", theorists start from a descriptive definition that outlines the provisional boundaries around the set of phenomena commonly called emotions. They do this by listing typical and apparent properties of emotions – such as that they have mental and bodily aspects with a certain heat, automaticity, urgency, and irrational flavor – and/or by listing typical emotion types – such as anger, fear, sadness, and joy. Next, they seek explanations that provide a unique common ground for these phenomena (i.e., fruitfulness criterion), while also accounting for the listed properties and emotion types (i.e., similarity criterion). When successful, this approach leads to a scientific definition of emotions. Most psychological emotion theories achieve this by combining constitutive explanations (specifying the components of emotions) with causal-mechanistic explanations (specifying causal factors and mechanisms). The most consequential diversity within the emotion-theoretical landscape lies in their causal-mechanistic explanations, which can further be organized using three forks. The first fork separates personal from social theories; the second splits personal theories into behavior-oriented and feeling-oriented ones; the third divides behavior-oriented theories into reactive and proactive ones. Reactive theories include evolutionary theories, network theories, and appraisal theories. Proactive theories are goal-directed theories. Feeling-oriented theories are psychological constructionist theories. Social theories can be seen as extensions of each of the other personal theories. The causal-mechanistic explanations offered by some of these theories have led to scientific definitions of emotions. Reactive theories define emotions as mental and physical changes caused by fixed cognition-motivation links: innate stimulus-response representations in evolutionary theories, learned stimulus-response representations in network theories, and abstract stimulus-response representations in appraisal theories. Some psychological constructionist theories define emotions as emotion-specific feelings caused by the categorization of bodily feelings. While a strict application of these scientific definitions holds promise for fruitfulness, the heterogeneity of the descriptive set poses challenges to similarity with this set. To solve this tension, some theorists favor fruitfulness over similarity. They do this by trimming the set of emotions, splitting it, or eliminating it – as seen in some goal-directed and some psychological constructionist theories. Others favor similarity over fruitfulness by adopting a cluster-type definition.

Keywords: emotion, definition, explanation, evolutionary, network, appraisal, goal-directed, constructionist

Introduction

The question "What is an emotion?" can be understood in two different ways: "What do people refer to when using the term emotion?" and "What is the scientific meaning of the term emotion?" Answers to the first question are known as descriptive definitions, those to the second question as scientific definitions (Widen & Russell, 2010). The two types of definitions can be situated in two different stages in the development of emotion theories (Moors, 2022). This development goes somewhat like this: The first stage is the articulation of a descriptive or working definition, which offers a provisional demarcation of the set of emotions (i.e., explanandum or to be explained phenomenon) from other sets, either by listing members or subsets of the set (i.e., emotion types) or properties of these members. The second stage is the proposal of an explanation in which emotions (the explanandum) are linked to an explaining entity (i.e., explanans). Two common types of explanations of emotions are constitutive ones, which specify the components that make up an emotion, and causal-mechanistic ones, which specify the causes of emotions and/or the mechanisms that mediate between these causes and these emotions. The third stage examines the proposed explanations in empirical research. If these explanations are validated and consensus is reached, a fourth stage follows, which is the articulation of a scientific definition. This definition is based on the explanations proposed in the second stage. Ideally, these explanations provide a deep common ground among the members of the set, allowing for the specification of necessary and sufficient conditions for determining membership from this point onward. Once a scientific definition of emotions is in place, other questions about emotions can be tackled in a more precise way, such as what the influence of emotions is on other phenomena (e.g., judgments, performance, well-being) or what happens when emotions are out of proportion (e.g., in psychopathology).

To make this way of looking at theory development more intuitive, consider the toy example of gold. Here, the working definition of the first stage specifies that gold can be provisionally demarcated with properties such as that it is shiny, yellow, dense, malleable, and ductile. A constitutive explanation in the second stage specifies that gold is made up of the atom Au. After empirical validation in the third stage, this constitutive explanation provides the basis for the scientific definition of gold. From now on, the descriptive definition of gold as shiny, yellow, dense and so on can be replaced with the scientific definition of gold as Au. This scientific definition allows diagnosing whether something is gold and thereby addressing all kinds of questions about gold in a more precise way, such as how gold participates in chemical bonding, or what the oxidation states of gold are.

While the scientific definition of gold is straightforward in that the constitutive explanation yields one necessary condition that is also sufficient (i.e., the presence of the atom Au), the common ground of other phenomena is often a compound of elements provided by constitutive and/or causal-mechanistic explanations. One example is the phenomenon of water. The constitutive explanation – H_2O – yields three necessary conditions that are together sufficient: two hydrogen atoms and 1 oxygen atom, bound together in a specific way. A more complex example is the phenomenon of a hangover. The constitutive explanation yields a number of (more or less) necessary conditions such as a headache, a dry mouth, and nausea. However, what is missing to demarcate hangovers from other phenomena, such as the flu, is the *cause* of these components and the mediating *mechanism*: excessive drinking and the ensuing metabolic changes in the case of hangovers, and a virus and the ensuing viral infection cycle and immune responses in the case of the flu. In other words, the causal-mechanistic explanation adds a crucial necessary condition to the set of conditions, which is now sufficient for demarcating hangovers from the flu. As will be clarified in later sections, most emotion theories in psychology consider emotions to be more like hangovers, recognizing that a constitutive explanation alone is not sufficient to demarcate emotions from non-emotional phenomena and that a causal-mechanistic story is required.

Despite differences in complexity, the scientific definitions of gold, water, and hangovers are relatively straightforward because their original descriptive sets are homogeneous enough to find a deep common ground among their members. Such a deep common ground increases confidence that the scientific set will turn out to contribute to scientific progress, a criterion called *fruitfulness*. An important aspect of fruitfulness is that the set allows for scientific extrapolation – allowing discoveries to be generalized across the members of the set (Griffiths 2004; Scarantino 2012). For instance, if one piece of gold is chemically inert, the same will apply to other pieces of the same purity; if one sample of water boils at 100 degrees, this will also hold for other samples (all else being equal); and if one instance of a hangover can be eased by taking fluids and nutrients, this will also be likely for other instances.

Many descriptive sets from natural language are heterogeneous, however. Think of games, art, mental illnesses, species like fish and lemons, and gemstones like jadestones. When a set is heterogeneous, it will be more difficult to find a single common ground for all its members, thereby jeopardizing fruitfulness. Yet fruitfulness can still be obtained if one is willing to relinquish the similarity of the scientific set with the original descriptive set. Conversely, one may also decide to relinquish fruitfulness to retain maximal similarity. In

other words, when a descriptive set is heterogeneous, a trade-off between fruitfulness and similarity is inevitable (Carnap 1947; Swartz 1997).

Prioritizing *fruitfulness over similarity* can be done in at least three ways. A first option is to narrow down the original, descriptive set. For instance, the descriptive set of fish, which houses both cold-blooded vertebrates with gills (e.g., sharks and goldfish) and warm-blooded ones without gills (e.g., dolphins and whales) has been regimented by discarding the latter ones. A second option is to split the original set into two (or more) novel sets. For instance, the descriptive set of jadestones, which originally comprised stones composed of two different minerals – jadeite and nephrite – has been split into two separate sets. A third, more radical, option is to eliminate the original set altogether and replace it with other, more fruitful sets. For instance, the descriptive set of air, characterized as transparent, odorless gas that fills the atmosphere and is inhaled, cannot be captured by a common constitutive explanation (it contains varying levels of oxygen, nitrogen, and carbonite, which are not bound together), and therefore has not attained scientific status. As such, it has been replaced with sets for each of its components.[i]

Prioritizing *similarity over fruitfulness* can be done by abandoning the quest for a classic definition in which a common denominator is proposed in the form of necessary conditions that are together sufficient, and by replacing it with a cluster-type definition, in which necessary conditions are treated as typical conditions instead (Boyd, 2010; Searle, 1958; Wittgenstein, 1953). For instance, since not all lemons are yellow (some are green), oval (some are round), and acid (some are bitter), each of these features is better conceived of as typical rather than necessary. Note that cluster-type definitions can be considered as disjunctions of sets of jointly sufficient conditions (Longworth & Scarantino, 2010). Thus, lemons can be defined as "yellow, oval and acid" or "yellow, round, and acid", or "green, oval, and bitter", and so on. Although cluster-type definitions succeed at maximizing similarity (they keep green, round, and bitter lemons on board), they come at a cost of fruitfulness (some discoveries about acid lemons may not extrapolate to bitter ones). This is why classic definitions remain researchers' first choice and why cluster-type definitions are only turned to when classic definitions are out of reach.

It may finally be noted that the causal-mechanistic explanations put forward by emotion theories not only serve to *demarcate* the set of emotions as a whole from other sets (captured in a scientific definition in *intensional* format; Carnap, 1947), but also to *partition* the set internally into subsets (corresponding to a scientific definition in *divisio* format; Ierodiakonou, 1993[ii]). To make this more intuitive, take the example of an allergic reaction.

This can be defined as symptoms that involve swelling of organs (constitutive explanation), which are caused by a foreign substance called an allergen (causal explanation), mediated by a reaction of the immune system (mechanism). The set of allergies can be internally organized based on types of allergens (e.g., pollen, bee venom, types of foods) but also based on types of immune reactions (e.g., IgE-mediated reactions and non-IgE-mediated reactions). The first way to partition is based on types of causes, the second on types of mechanisms.

In the upcoming sections, these insights about definitions are applied to the set of emotions. First, the section "The Descriptive Definition of Emotion" provides a discussion of the descriptive definition of emotion. Next, the section 'Scientific Definitions of Emotion" elaborates on the scientific definitions of emotion proposed by various theoretical traditions, based on their favorite explanations. It considers not only how these explanations help demarcate the set of emotions from other sets, but also how they help partition the set into subsets. Finally, the section "Adequacy of Scientific Definitions" takes stock and evaluates the adequacy of these definitions in terms of fruitfulness and similarity, essentially asking whether emotions are analogous to gold, water, hangovers, jade, lemons, or air.

The Descriptive Definition of Emotion

The descriptive definition of emotion draws the provisional contours of the set of emotions, based on the way in which lay people have used the term in daily life. As mentioned, such a definition may articulate properties of emotions capturing intuitions about what they have in common and what makes them unique (i.e., intensional format) or it may come in the form of a list of prototypical emotion types (i.e., divisio format). So when asked "What is an emotion?", people can either list a number of properties or a number of emotion types. Although descriptive definitions draw on the lay usage of the term emotion, they do set the agenda for scientific emotion theories. The emotion properties and types that are part of ordinary understanding of emotions fix the desiderata that emotion theories should fulfill – the stuff about emotions that require an explanation. A brief look at these emotion properties and types is in order, starting with the types.

Emotion types for which there is fair consensus are fear, anger, sadness, joy, disgust, and surprise (and shades of these emotions such as anxiety, terror, irritation, frustration, elation, and shock). Many theorists extend their lists with shame, guilt, regret, envy, pride, gratitude, indignation, jealousy, awe, admiration, and so on.

Before turning to the properties, it is worth clarifying that because descriptive definitions are provisional demarcations, the status of these properties is that they are (a) typical rather than necessary (they may not be shared by all members of the descriptive set)

and (b) apparent rather than real (they remain open to revision). Typical and apparent properties of emotions about which there is reasonable consensus have to do with the *where*, *when*, *what*, *how*, and *why* of emotions. Emotions are located in the individual and they can take place in human adults as well as in human infants and non-human animals (where). They are elicited when events of major importance occur (when). These events comprise both (a) physical events in the environment, such as the loss of a friend, the encounter with a vicious dog, and an unjust treatment; and (b) mental events (i.e., types of mental representations) such as memories, expectations, and imaginings.

Emotions themselves are also made up of mental and bodily aspects (what). The mental aspect of emotions is often captured by the statement that they are about something or have object-directedness. Thus, one can be sad *about* the loss of a friend, scared *of* a vicious dog, and angry *about* an unjust treatment. The bodily aspect of emotions refers to the observation that they are accompanied by physiological responses such as increased heart rate, blood pressure, and blushing, as well as subtle and coarse forms of overt behavior. The subtle forms refer to facial, vocal, and gestural expressions, such as a happy or scowling face, shouting or whining, and cringing or expanding the chest. The coarse forms encompass musculoskeletal behavior or actions, or at least the tendency to engage in these actions – called action tendencies (Frijda, 1986). Examples (on diverse levels of abstraction) of coarse behaviors are approach, avoidance, fight, flight, freezing, and repair. To demarcate emotions from other mental and bodily phenomena, they have been characterized as “hot” and the non-emotional stuff as “cold”. The heat of emotions has a quantitative aspect, which refers to their intense character, and a qualitative aspect, which refers to their positive or negative valence.

Emotions arise in an automatic fashion (how), which means that they can arise quickly, when attention is scarce or directed elsewhere, without intention, and/or against the individual’s will. They are also disruptive, meaning that they clamor priority over ordinary goal pursuit – a feature coined as “control precedence” (Frijda, 1986).

There is less consensus about whether emotions have a function or are merely disruptive (why). Do they make individuals think more or less clearly? And do they make them act in a more or less adaptive way? Theorists have separated the question of whether emotions are appropriate or inappropriate in a certain situation (i.e., theoretical rationality) from the questions whether they improve or bias thinking and behavior (i.e., practical rationality). Public opinion, on the other hand, undeniably contrasts emotions with ratio. Therefore, irrationality counts as a typical and apparent property of emotions that begs for an explanation.

Scientific Definitions of Emotions

Once the descriptive definition of emotions is in place and the descriptive set of emotions is provisionally demarcated from other sets, theorists develop explanations in the hope that these will capture the deep common ground between the members of the set. In other words, these explanations should yield the criteria or conditions for determining membership in the ultimate, scientific set of emotions. This section discusses the various attempts that theorists have undertaken in this regard.

The criteria put forward by emotion theories are based on constitutive as well as causal-mechanistic explanations of emotions, similar to what was discussed for hangovers. Hangovers are seen as composed of nausea, a headache, and a dry mouth (constitutive explanation), caused by excessive drinking that sets in motion a chain of internal metabolic reactions (causal-mechanistic explanation). Likewise, emotions are seen as episodes with one or more components (constitutive explanation), caused by internal and external events that set in motion some kind of mechanism (causal-mechanistic explanation). Unlike for hangovers, however, there is no consensus about the components and causal mechanisms of emotions. The remainder of this section zeroes in on the divergence that exists in the literature for constitutive and causal-mechanistic explanations separately.

Constitutive Explanations

The constitutive explanations of emotions range from narrow to broad. On the narrow end of the spectrum are philosophical and older psychological theories in which emotions are identified with a single component. Some of them take emotions to be special types of feelings in line with lay understanding (e.g., James, 1890; LeDoux, 2012; Reisenzein, 2012), setting apart feelings of fear and anger, for instance, from feelings of cold, hunger, and physical pain. Others have treated emotions as special types of cognitions or evaluations (understood as judgments, Solomon, 1993; quasi-judgments, Armon-Jones, 1991; or perceptions, Tappolet, 2016). For instance, fear is the evaluation of an event as dangerous and anger the evaluation of an event as offensive. Still others consider emotions as special types of motivations (e.g., Roseman, 2013; Shand, 1914). Here, fear is considered as the tendency to flee and anger as the tendency to fight, for example. Finally, some theorists consider emotions to be special types of observable bodily responses (e.g., Watson, 1919). Thus, fear and anger are the patterns of physiological and motor responses involved in fleeing and fighting behavior.

Dissatisfied with the strong emphasis that these theorists place on a single component, later psychological theories have opted for a broader constitutive explanation, including two

or more components, with the most inclusive proposal being that emotions are made up of all the components just mentioned (i.e., feelings, cognitions, motivations, physiological responses, and behavior). Note that the latter theorists tend to think of emotions as episodes rather than states, often replacing talk of "emotions" with talk of "emotional episodes" (Clore & Ortony, 2000; Frijda & Parrott, 2011; Scherer, 2009).

Causal-mechanistic Explanations

Turning to the causal-mechanistic explanations, a first observation is that theories with a narrow constitutive explanation (identifying emotion with fewer components) often treat entities that others consider to be components as causes and consequences of emotion. For example, James (1890) identified emotions with feelings of bodily responses (constitutive explanation), which he saw as caused by stimuli that activate instincts and habits (causal-mechanistic explanation). Instincts and habits are understood as respectively innate and learned associations between stimulus representations and response representations ([S-R]). Stimulus representations are forms of cognition whereas response representations – which can be considered as action tendencies – are forms of motivation. In other words, James (1890) viewed the connection between the cognitive and motivational components as the causal mechanism of emotion.

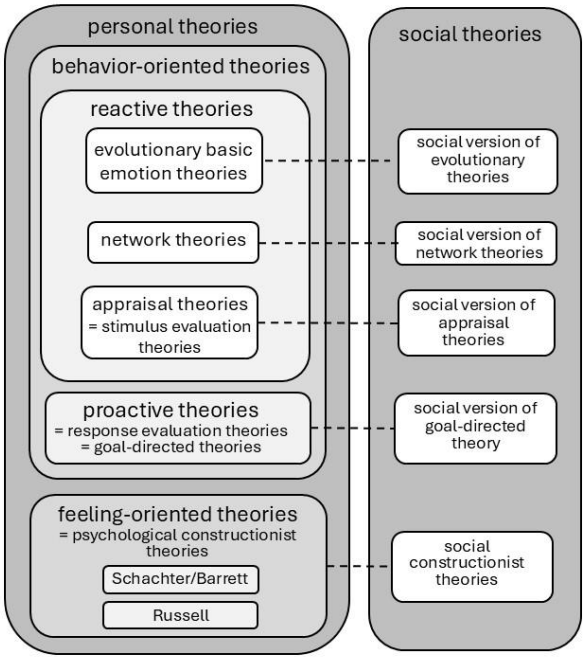
A second observation is that theories with a broad constitutive explanation (identifying emotion with more components) often do not propose a mechanism that precedes and literally causes the emotion but rather one that is part of the emotion but that can still cause the other components in the emotion. Upon close scrutiny, many of these theories also take the connection between the cognitive and motivational components to be this mechanism. This underlines that a theory's breadth of constitutive explanation is relatively independent of its causal-mechanistic explanation.

The prevalent causal-mechanistic explanations can be organized with the help of three forks (see Figure 1; see also Moors, 2024). The first fork splits theories into (a) personal theories, which situate their key mechanisms inside the individual, and (b) social theories, which focus on the social context or social mechanisms. The second fork splits the group of personal theories into (a) behavior-oriented theories, which prioritize explaining the behavior component of emotions, and (b) feeling-oriented theories, which prioritize explaining the feeling component of emotions. The second fork does not represent a sharp distinction given that behavior-oriented theories are not silent about feelings and feeling-oriented theories are not silent about behavior. The distinction may be more relevant historically than it is today, but in this way, it may help understand how the two traditions grew apart. The third fork splits

behavior-oriented theories into (a) reactive theories, which take fixed (innate or learned) connections between the cognitive and motivational components as the key mechanism of emotions and (b) proactive theories, which argue for a flexible connection between these two components.

Several families of emotion theories can be situated in the resulting four classes. Reactive theories encompass evolutionary basic emotion theories, network theories, and stimulus evaluation or appraisal theories. Proactive theories are response evaluation theories, better known as goal-directed or instrumental theories. Feeling-oriented theories include psychological constructionist theories. Social theories, finally, encompass social versions of personal theories, including social constructionist theories. The following section elaborates on the causal-mechanistic explanations of these four classes of theories and discusses the scientific definitions of emotions that can be derived from these explanations. As will be clarified, James' s (1890) seminal theory occupies a central position in the history of emotion theory, serving as an early source of inspiration for both behavior-oriented and feeling-oriented theories.

Figure 1. *Emotion theories in psychology*



Reactive Theories

In reactive theories, internal or external events trigger an association between a stimulus representation (cognitive component) and a response representation or action tendency (motivational component). This association is taken to be the backbone of all emotions. An early version of this idea can be found in James's (1890) theory. According to him, stimuli activate instincts (innate S-R associations) and habits (learned S-R associations), which in turn produce overt bodily responses (physiological and motor responses). The conscious experience of these bodily responses is what James (1890) calls emotions. His scientific definition of emotions is thus that emotions are bodily feelings generated by instincts or habits.

James's (1890) scientific set of emotions appears to be overinclusive in that it keeps on board bodily feelings that arise from instincts and habits that are not included in the descriptive set of emotions. Examples are the bodily feelings arising from the foraging instinct in animals and the grasp reflex in babies. Another objection against James's (1890) proposal was that in his time, empirical research showed that bodily responses produced by instincts and habits were undifferentiated, or at least that they did not follow the fault lines presented by distinct emotion types from natural language.

Different subsequent theories have provided different solutions to the problems presumably plaguing James's (1890) theory. The remainder of this section outlines the solutions proposed by three theories that qualify as reactive ones: evolutionary theories, network theories, and appraisal theories.

Evolutionary Theories.

To solve the problem that James' (1890) scientific set was overinclusive, evolutionary theories proposed that emotions stem from instincts alone (not habits) and only from a particular type of emotional instincts, which go by the name of affect programs. However, the boundary between affect programs and other instincts was still drawn artificially, with affect programs matching a handful of emotion types from natural language, called basic emotions: fear, anger, joy, sadness, surprise, and disgust (Ekman, 1999; Ekman & Cordaro, 2011).[iii]

The empirical efforts of evolutionary theorists have concentrated on delivering direct as well as indirect evidence for the existence of affect programs (e.g., see reviews by Ekman, 1992; Kreibig, 2010; Keltner et al., 2019; Levenson, 2014; Shiota, 2024). Direct evidence is evidence that each basic emotion is subserved by a distinct neural substrate. While early research has tried to identify specific neural *circuits* (i.e., interconnected brain regions), later research has tried to identify specific *patterns of whole brain activity* (see Shiota, 2024). To

date, however, the evidence is not convincing and remains open to alternative interpretations (Lindquist et al., 2012; Moors, 2022).

Indirect evidence is evidence that each basic emotion is characterized by a distinct facial expression and/or a distinct pattern of physiological responses (and that these are universal across cultures). In searching for this evidence, evolutionary theorists thus renewed the efforts that had been unsuccessful in James's (1890) time. Although the evidence seemed encouraging at first, it was questioned on methodological as well as logical grounds (Barrett, 2006a; Mauss et al., 2024; Ortony & Turner, 1990; Russell, 2003; see also Moors, 2022, Chapter 4).

The lack of convincing evidence has not stopped the advocates of this theory from pushing forward a scientific definition of emotion saying that an emotion is the sum of mental and bodily components that arise from activating an affect program. To partition the set of emotions, moreover, evolutionary theories propose that different affect programs ground different basic emotions.

Network Theories

Network theories of emotion, at least in their original form (Bower, 1981; Lang,1994; Leventhal, 1984), have accepted the existence of a limited set of basic emotions, but they have emphasized that the organism forms representations of the components of these emotions, which are encoded in memory where they are stringed together into emotion networks. In addition, they have suggested that each emotion network can become elaborated through learning principles. Indeed, any event, response, or verbal label that co-occurs with the basic emotion can be added to the original network and become part of it. To be fair, evolutionary theories already acknowledged that learning expands the range of stimuli that can activate basic emotions as well as the range of responses that organisms have at their disposal. But while evolutionary theories consider only the innate stuff to be emotional – the hard liquor so to speak (Prinz, 2004) – network theories also include the learned stuff as genuinely, if not more, emotional.

Modern versions of network theories are no longer committed to the idea that learned emotion networks should be centered around affect programs. They highlight instead that the clustering of components in a network solely relies on the interactions among these components (Colombetti, 2009; Lange et al., 2020; Lewis, 2005; Lewis & Liu, 2011; see Moors, 2022, p. 153; for a hybrid view, see Wood & Coan, 2023). According to them, emotion networks can but do not have to match up with the basic emotions proposed by evolutionary theories.

The scientific definition of emotion flowing from both original and modern network theories is that an emotion is the sum of mental and bodily components that arise from activating an emotion network in memory. This mechanism also helps partition the set of emotions into subsets in that each emotion is aligned to a separate emotion network.

Appraisal Theories.

A final strand of emotion theories that fit in the category of reactive theories are stimulus evaluation theories, better known as appraisal theories (Arnold, 1960; Lazarus, 1991; Roseman, 2013; Scherer, 2009; see Moors et al., 2013). Here too, associations between stimulus representations and response representations or action tendencies (S-R) form the backbone of emotions. Important to add, however, is that the stimulus representations that are part of these associations are not concrete stimulus representations, such as of an approaching car, a shouting boss, the death of a friend, or a successful application. Instead, they are abstract stimulus representations that result from a process of stimulus evaluation or appraisal. Examples are appraisals of danger, obstacle, offense, irrevocable loss, and success. Appraisal theories typically decompose these "molar" appraisals into patterns of "molecular" values on a number of appraisal variables or criteria such as the degree of goal relevance, goal in/congruence, un/expectedness, ease/difficulty to control, internally/externally caused, and so on. To illustrate, "danger" has been decomposed into the pattern "goal-relevant, goal-incongruent, unexpected, difficult to control, and externally caused", "offense" into the pattern "goal-relevant, goal-incongruent, easy to control, and externally caused", "irrevocable loss" into the pattern "goal-relevant, goal-incongruent, impossible to control, and externally caused", and success into the pattern "goal-relevant and goal-congruent".

Original formulations of appraisal theories (e.g., Lazarus, 1990) remain close to evolutionary theories in that they take the associations between molar appraisals and action tendencies to be innate and hardwired. Modern versions (e.g., Scherer, 2009; Scherer & Moors, 2019) allow each of the molecular appraisal values to separately determine one aspect of the initial action tendency and the ensuing components. For instance, goal-relevance determines the intensity of the action tendency, goal-in/congruence determines its direction (toward/away), and ease/difficulty to control determines the direction of adaptation (if easy/difficult to control, the stimulus/organism is adapted/adapts to the organism/stimulus). Nevertheless, the associations between molecular appraisal values and each of these action tendency aspects remain a rigid affair, whether innately given or based on learning. This is why appraisal theories, no matter how sophisticated they are fleshed out, ultimately remain reactive theories.

The scientific definition of emotion proposed by original and modern appraisal theories is that an emotion is the sum of mental and bodily components that arise from the appraisal of a stimulus as goal-relevant.[iv] Since goal-relevance is best understood as a dimensional variable, it follows that appraisal theory's scientific definition can only provide a gradual distinction between emotional and non-emotional phenomena.

The appraisal process is also used as the basis for partitioning of the set of emotions. Different appraisal patterns produce different emotion types. In original appraisal theories, the resulting emotions correspond to the basic emotions listed by evolutionary theories. Modern appraisal theories reason that because the number of appraisal patterns is infinite, the number of emotions should be infinite as well. This variety is not neatly captured into basic emotion subsets. Instead, this variety is better described using a matrix formed by the various appraisal variables in which individual emotions can be placed.

The proposal of appraisal theories that incoming stimuli are first given meaning through an appraisal process before they activate action tendencies has been taken over by modern versions of evolutionary theories (so that these become indistinguishable from original appraisal theories) as well as by network theories (which take appraisal to be one node within the network). Again, no matter how sophisticated the extraction process that occurs when moving from stimulus to stimulus representation (i.e., with or without appraisal), the translation process that occurs when moving from stimulus representation to response representation or action tendency remains a stimulus-driven process in all reactive theories.

All reactive theories discussed so far are dual-process theories in which stimulus-driven processes involved in emotions are contrasted with goal-directed processes involved in instrumental behavior and emotion regulation. The hallmark of a goal-directed process is that (overt or covert) behavior is the result of the weighing of the expected utility of the available action options in the person's action repertoire. The notion of expected utility can be unpacked as the product of the value of the outcomes of the action and the expectancy (i.e., subjective probability) that the action will indeed lead to these outcomes.

Instrumental behavior can be entirely separate from the emotion, but it is often seen as part of the implementation of the initial, emotional action tendency into concrete action tendencies and behavior (i.e., *planning*). Thus, after an appraisal of danger has activated the abstract action tendency to defend oneself, a goal-directed process may take over to translate this tendency into the more concrete tendency to flee, fight, or hide depending on which of those actions have a higher expected utility in the current context. Goal-directed processes are also assumed to be involved in *emotion regulation* (Gross, 2015). Thus, after an appraisal of

offense has activated the abstract action tendency to aggress, a goal-directed process may take over to alter this tendency into the tendency to negotiate if aggression goes against the person's goal to be safe and avoid retaliation.

The interplay between stimulus-driven and goal-directed processes, moreover, is understood as following a default-interventionist architecture (Moors et al., 2017). Stimulus-driven processes involved in emotion are supposed to be automatic, which is why they are triggered by default. Goal-directed processes involved in planning and regulation are supposed to be less automatic and to intervene when operating conditions are ample, such as when there is sufficient time, attention, and motivation to do so. The goal-directed theory, to be discussed in the next section, breaks with this architecture and proposes instead that not only planning and regulation but also emotions themselves rely on goal-directed processes.

Proactive Theories

Proactive theories encompass goal-directed or instrumental emotion theories (e.g., Broekens et al., 2015; Rolls, 2005) of which Moors's (2017a, 2017b, 2022; Moors et al., 2017) version is an articulate example. This theory challenges the long-standing idea espoused by reactive theories that emotions are based on innate or learned stimulus-driven processes, which can be complemented or counteracted by goal-directed processes.

Rather than relegating goal-directed processes to planning and regulation, the goal-directed theory places goal-directed processes at the heart of emotion generation. The theory's mechanistic explanation for most of the phenomena that people call emotions (i.e., members of the descriptive set of emotions) goes roughly as follows. Organisms are constantly trying to fulfill their goals by engaging in goal-directed cycles. A typical goal-directed cycle starts with the detection of a discrepancy between a current or anticipated stimulus and a first goal, which activates a second goal to reduce this discrepancy. To achieve this second goal, the organism can select between one of three broad strategies: (a) assimilation, which is to engage in behavior to change the stimulus in the direction of the first goal, (b) accommodation, which is to change the first goal in the direction of the stimulus, and (c) immunization, which is to reinterpret the stimulus so that it appears as less or no longer discrepant with the first goal. If the organism selects assimilation, this also entails selecting between different action options in the behavior repertoire. Once such an action option is chosen, the corresponding action tendency – which is a third goal – is activated. This is subsequently implemented in actual behavior, leading to an actual outcome. The outcome is the new stimulus input, which is fed back to the system where it initiates a new cycle. This continues until there is no discrepancy left, or until the goal-directed cycle at the service of another, more important goal takes over.

To illustrate, suppose Sam receives a harsh remark from his boss (actual stimulus), which is discrepant with his goal to be treated with respect (first goal). To reduce this discrepancy (second goal), Sam may choose between several behaviors from his repertoire (i.e., assimilation): act aggressively (shout and slam with the door), act in a submissive way (shrink posture and apologize), or act in a reconciling manner (explain calmly that he feels disrespected). If Sam chooses the aggressive option, this will activate an aggressive action tendency (third goal), and this may be implemented in overt behavior and produce an actual outcome. But Sam could also give up his goal of getting respect from his boss and start looking for another place to work (i.e., accommodation), or he could reinterpret his boss's remark as honest, well-intended feedback (i.e., immunization).

Both the selection of broad strategies (assimilation, accommodation, and immunization), and of action options in the case of assimilation, are done by weighing up the expected utilities of the strategies or action options, given the current situation. As mentioned earlier, the expected utility of a single action option depends on the value of the outcome of the action and the expectancy that the action will indeed produce the outcome. But since every organism has multiple (sometimes conflicting) goals, action options may be conducive to some goals (i.e., benefits) but not others (i.e., costs).

All components typically identified in emotional episodes are present in this cycle. The information processing involved in discrepancy detection and strategy/action selection corresponds to the cognitive component. The action tendency corresponds to the motivational component. Actual, overt behavior corresponds to the motor component, and the mobilization of physiological activity to prepare for and support this behavior corresponds to the somatic component. Certain aspects from all the steps in the goal-directed cycle may surface into consciousness and the sum of all these aspects corresponds to the feeling component. These feelings can but do not have to be labelled as emotional or as a specific emotion.

When Sam, after receiving a harsh remark from his boss, selects aggressive behavior, he may be more likely to consider the episode to be emotional (and label his feelings as angry) than when it results in negotiating behavior. Crucially, however, both behaviors are produced by the same goal-directed process. This underscores the theory's core claim that goal-directed processes play an equal role in emotional and non-emotional episodes.

The goal-directed theory does not a priori deny the existence of innate or learned stimulus-driven processes but considers their role to be negligible. Organisms may well come into the world with certain instincts (i.e., innate stimulus-driven processes), but grounding emotions in instincts faces two problems: First, the set of instincts shows only partial overlap

with the set of phenomena typically called emotions. Second, once an instinct has produced behavior, it will be followed by an outcome. Once this outcome gets encoded in the system, the organism will from then on be able to take into account this outcome. Thus, once enacted, instincts may quickly become overwritten by goal-directed processes.

Organisms may also acquire habits (i.e., learned stimulus-driven processes), based on the idea that organisms may form (fleeting) associations between any two or more representations. The goal-directed theory does not accept the route in which habits are the end point of overtraining goal-directed processes. This is because the theory proposes that overtraining goal-directed processes leads to stronger rather than weaker goal-directed processes (Buabang et al., 2023; Moors et al., 2017).

Because the goal-directed theory accepts the existence of innate or learned stimulus-driven processes next to goal-directed ones, it still qualifies as a dual-process theory. The architecture describing the interplay between the two processes, however, is not default-interventionist, as in reactive theories, but parallel-competitive. In particular, the theory assumes that not only stimulus-driven processes but also goal-directed ones can be automatic. This means that they can both operate when time, attention, and motivation are scarce. As a result, the two types of processes will often operate in parallel so that they enter in competition with each other if each dictates a different action tendency. If that happens, the theory further hypothesizes that the goal-directed process will often win the competition. Thus, the theory predicts that goal-directed processes will be the main determinant of behavior – regardless of whether this behavior is seen as part of an emotional or a non-emotional episode.

Given that the causal-mechanistic explanation of the phenomena called emotions is not unique to emotions but shared by non-emotional forms of goal pursuit, this explanation does not provide a criterion for the demarcation of the set of emotions. The goal-directed theory nevertheless recognizes that there may be descriptive differences between episodes that people do and do not call emotional. A key difference is that the goals at stake in emotional episodes are often of higher value, resulting in more intense and urgent action tendencies and a more vigorous and acute mobilization of physiological activity required to implement these action tendencies in overt behavior. This corresponds to the goal relevance criterion of appraisal theories. However, unlike appraisal theories, the goal-directed theory does not consider the criterion of goal value or goal relevance to be deep enough to grant emotions scientific status.

Given that the goal-directed theory does not consider the set of emotions as a scientific set in its own right, there is no point in asking what the best way of partitioning this set is or whether carving it up into vernacular emotion subsets is sound. Nonetheless, it may be virtuous for any theory to be able to make sense of such subsets. This amounts to answering the question of how people come to label an episode as anger, fear, sadness, joy, and so on. Labeling may be based on different elements in the goal-directed cycle: the type of discrepancy (current or anticipated), the goal (i.e., first goal) at stake, or the action tendency (i.e., third goal) that was activated. Let me illustrate this for anger and fear. For anger, let's return to Sam who, after receiving a harsh remark from his boss, is thorn between the tendency to aggress and the tendency to withdraw. Sam may label this episode as anger because he encountered a current discrepancy with some goal (i.e., an obstacle), because the discrepancy was with the goal to be respected (i.e., the obstacle was in fact an offense), and/or because he felt a tendency to be aggressive. For fear, the illustration can be tweaked as follows. Suppose Sam enters the office of his boss, who frowns at him, signaling that she is about to give Sam a harsh remark (i.e., anticipated stimulus). This creates an anticipated discrepancy with the goal to be respected or even the goal to keep his job, which Sam may try to reduce by selecting one among several action options: avoidance (run out the office), submission (apologizing preemptively), or appeasement (by offering to take on extra work). Again, Sam may label this episode as fear because there was an anticipated discrepancy with some goal (i.e., a threat), because the discrepancy was with the goal to keep his job (i.e., the threat concerned security of safety), and/or because he had selected the tendency to avoid, be submissive, or appease.

Feeling-oriented Theories

As indicated earlier, James's (1890) theory can be considered as the precursor not only of behavior-oriented theories but also of feeling-oriented theories, in particular psychological constructionist theories. As may be recalled, James (1890) proposed that emotions are feelings of bodily responses caused by instincts or habits. This proposal was deemed unsatisfactory because (a) the set of emotions carved out by all instincts and habits seems overinclusive and (b) cannot ground the vernacular emotion types – as suggested by the lack of evidence that specific emotions were characterized by specific patterns of bodily responses. Evolutionary theorists proposed to solve these problems by (a) artificially reducing the set of emotions to a handful of instincts (called affect programs) that match with vernacular emotion types – the basic emotions – and (b) continuing the quest for distinct patterns of bodily responses for these basic emotions.

Psychological constructionist theories propose a different solution. They reject the idea that emotions are generated by dedicated mechanisms. Thus, they maintain that bodily responses are not emotion-specific, and so neither are the feelings based on these responses. To turn these bodily feelings into emotion-specific feelings, they need to be categorized as such. Thus, Schachter (1964) proposed a two-factor theory in which diffuse bodily feelings (Factor 1) are attributed to an external situation, thereby allowing these bodily feelings to be categorized in terms of a specific emotion (Factor 2). For instance, Sam experiences the same bodily responses when encountering a vicious dog and when hearing about the betrayal of a colleague, but these responses turn into fear and anger only after they have been attributed to their presumed causes. Building on these seminal ideas, Barrett (2006, 2012, 2014, Barrett & Lida, 2024) proposed a construction process in which emotion-specific feelings arise when sensations from inside (called core affect) and outside the body (i.e., external stimuli) are integrated with learned knowledge about emotion categories, referred to as emotion concepts. An emotion concept (e.g., anger) is a set of situation-specific scripts (e.g., anger in traffic, insult, or injustice), each encoding information about a situation, a behavior, and a sensory outcome in a multimodal, embodied format. In early formulations of the theory (Barrett, 2006, 2012), internal or external inputs can trigger full script activation via pattern completion, leading to partial re-experiencing and re-enactment (much in line with network theories). Later, emotion concepts were reframed as predictions, which are compared to sensory input to settle on the best script (Barrett, 2017). Zooming out from the mechanistic details, Barrett (2006) advocated for a strong form of emergence, where the whole (the emotion) is more than the sum of the parts (the components).

Thus, Schachter (1964) and Barrett (2012) put forward a scientific definition in which emotions are bodily feelings categorized or constructed as emotions or as specific emotion types. The categorization of an experience as an emotion not only places boundaries around the scientific set of emotions; it also allows partitioning this set into the vernacular subsets that many researchers are after.

A different version of psychological constructionism propelled by Russell (2003, 2009, 2012, 2014) states that the phenomena that people call emotions correspond to episodes, which contain components such as information processing (cognitive component), action tendencies (i.e., motivational component), physiological responses (somatic component), behavior (motor component), and feelings (feeling component). The components do not stem from a pre-set mechanism – neither an innate program nor a learned network –

but each component is undergirded by its own set of mechanisms. Construction here means that the components are assembled in the moment.

Taking a closer look at the feeling component in Russell's (2014) theory, one can still recognize the two-factor idea of Schachter (1964). The feeling component comprises diffuse bodily feelings (i.e., core affect) as well as emotion-specific feelings (i.e., emotional meta-experience). Emotion-specific feelings arise when diffuse bodily feelings, and other components present in the episode, are categorized in terms of a specific emotion. This categorization process is based on the resemblance of the current configuration of components with a stored mental script for a particular emotion. Mental scripts for emotions are composed of (representations of) prototypical components (e.g., an anger episode is composed of an obstacle or offense and an aggressive tendency; a fear episode is composed of a threat and an avoidance tendency). In contrast to Schachter (1964) and Barrett (2006), Russell (2014) does not consider the result of the categorization process (i.e., emotional meta-experience) to be an emotion but rather the *self-ascription* of an emotion. Loyal to the idea that all mechanisms at play in episodes that people call emotions are general-purpose and not dedicated to emotions, this should equally apply to the categorization process. Thus, the categorization process can explain how people come to *think* they have an emotion, but it does not buy the set of emotions scientific status. The set of emotions remains a descriptive set or "folk set".

Social Theories

The previously discussed theories count as "personal" theories because they situate emotions and their mechanisms within the individual. Although they acknowledge that the majority of emotions occur in a social context and are triggered by social stimuli (e.g., the behavior of others), they usually do not elaborate on this much. Some theorists, however, have developed "social" extensions of these theories, collectively referred to here as "social theories". These social theories have explored the implications of the social context on emotions in much greater detail. By expanding the focus from a single individual to several individuals (i.e., dyads, groups, and cultures) and from a single emotional episode to multiple interlocking episodes, social theorists have even come up with distinct social mechanisms for emotions, that is, mechanisms that are not situated within but rather between individuals. Let us briefly consider some of these social theories (see Moors, 2022, Chapter 9).

Social Evolutionary Theories

One social extension of evolutionary theories asks about the evolved social functions of emotions in dyads, groups, and cultures (Keltner & Haidt, 2001; Keltner et al., 2006). In dyads, emotions help coordinate interactions between two people. In groups, emotions help

shape the roles of individuals within these groups as well as the identity of groups as a whole. In broader cultures, the expressions of emotions are regulated by sociocultural norms.

Another social extension of evolutionary theories focuses on the evolutionary origin of emotional expressions in communication (Shariff & Tracy, 2011; Sauter & Russell, 2025). Initially, certain facial displays developed as adaptations, like baring one's teeth to prepare for biting. Over time, these displays evolved into exaptations, serving new functions such as communication. For instance, baring one's teeth evolved into a signal of anger (but see Fridlund, 1994, 2017; Moors, 2022).

Social Network Theories

A notable social extension of network theories is the proposal of emotional contagion as a social mechanism of emotion causation (Hatfield et al., 1994). When a receiver mimics the emotional responses of a sender (e.g., a sender's scary face), somatosensory feedback of these responses induces a corresponding emotion (e.g., fear) in the receiver. This mechanism fits with the assumption in network theories that emotions can be activated not only via the stimulus side (e.g., by a dangerous event in the environment) but also via the response side (e.g., by artificially putting on a scared face).

Social Appraisal Theories

One social extension of appraisals theories asks whether appraisal criteria are weighted differently in different cultures and whether novel appraisal criteria can be discovered in non-western cultures. (e.g., Markus & Kitayama, 1991; Mesquita & Ellsworth, 2001). For instance, western (individualist) cultures may put more weight on whether a goal-incongruent event was internally/externally caused, which could result in more instances of anger compared to non-western (collectivist) cultures.

Another social extension is the proposal of social appraisal as a social mechanism of emotion causation (Manstead & Fischer, 2001). When a receiver infers how to appraise a stimulus based on the emotional responses of a sender (e.g., that a stimulus is offensive based on the sender's angry facial expression), this may lead to a corresponding (e.g., anger) or complementary emotion (e.g., fear).

A final social extension is the mechanism of distributed appraisal as a social mechanism of emotion causation (Parkinson, 2012). Here, the various appraisal checks are not made all at once within a single individual, but they are negotiated with other individuals across multiple interaction units. For instance, an event that is initially appraised by Sam as novel may subsequently, through the interaction with a colleague, be appraised as goal-

incongruent, and later alternatingly as easy vs. difficult to control. This may cause Sam's emotions to oscillate between anger and fear.

Social Goal-directed Theory

One social extension of the goal-directed theory (Moors, 2022) proposes to treat facial expressions not as direct read-outs of emotions but as forms of behavior that are instrumental for reaching a variety of goals (e.g., Fridlund, 2017; Moors, 2022; Parkinson, 2005). These goals may include communication goals, not just the goal to communicate (what one believes is) an emotion, but also the goal to communicate one's own intentions and expectations from others. These goals may also include non-communication goals, where the individual does not want to convey any message, but strives to influence other individuals directly (Parkinson, 2007). To illustrate these possibilities, a frowning face may not express anger but may be a goal-directed attempt to communicate that one (believes one) is angry, that one has the urge to punch, or that one had expected something else. It may also be a direct attempt to solicit consolation from others without conveying any symbolic message.

Another social extension of the goal-directed theory is the proposal of a distributed goal-directed process in which the weighing up of the expected utilities of action options does not occur in a single shot in a single individual but is spread out and adjusted across multiple interaction units with other individuals (Moors, 2022).

Social Constructionist Theories

Social extensions of psychological constructionist theories go by the name of social constructionist theories. One such extension emphasizes the role of culture in shaping the knowledge drawn upon in the categorization process responsible for transforming core affect into specific emotions (Averill, 2012; Mesquita & Parkinson, 2024[v]). The most striking evidence of this is the existence of emotion words that are unique to specific cultures (e.g., "amae" in Japan, "Schadenfreude" in Germany, "saudade" in Portugal). Another social extension proposes that emotion categorization is negotiated in social interactions rather than completed in the mind of the individual. This has led to the proposal that emotions are phenomena located *between* individuals rather than *within* them (Mesquita, 2022).

After having discussed these social theories, what are the implications for the scientific definitions of emotions? Social extensions of personal theories should generally propose the same scientific definitions as their personal counterparts and the same methods for partitioning this set. The only exception may be radical social constructionist theories, which argue that emotions are located between rather than within individuals, at least if this

statement is meant to be taken literally rather than rhetorically (Colombetti & Krueger, 2015; but see Griffiths & Scarantino, 2005; Léon et al., 2019).

Adequacy of scientific definitions

This final section evaluates the adequacy of the scientific definitions reviewed so far in terms of fruitfulness and similarity. To reconnect with the introduction section, it asks whether emotions are analogous to gold, water, hangovers, jade, lemons, or air.

It is safe to say that most theorists cherish the ideal to find a scientific intensional definition of emotion in which criteria for membership form an elegant set of necessary and sufficient conditions. As already indicated, most emotion theories propose a combination of criteria derived from their constitutive and causal-mechanistic explanations – analogous to the definition of hangovers. All the scientific intensional definitions discussed so far fall under this approach. Reactive theories define emotions as mental and physical changes caused by fixed cognition-motivation links. The precise nature of these links varies across the different reactive theories discussed. Indeed, in evolutionary theories, they are innate [S-R] links known as affect programs. In network theories, they are learned [S-R] links, with an entire network of additional learned links attached to them. And in appraisal theories, they are more abstract [S-R] links, with the "S" part corresponding to the output of an appraisal process. Some feeling-oriented theories, such as some versions of constructionist theories, likewise present a scientific definition analogous to hangovers. They define emotions as emotion-specific feelings caused by the categorization of bodily feelings.

As argued earlier, however, the adequacy of a scientific definition not only hinges on fruitfulness but also on similarity between the scientific set carved out by the scientific definition and the original, descriptive set. The heterogeneity of the descriptive set of emotions presents challenges to similarity.[vi] Due to this, the hard definitional criteria listed by the above theories have met with criticism in the form of counterexamples: members of the descriptive set that fall outside of the scientific set or, conversely, members of the scientific set that were not part of the descriptive set. A few examples. Fear of missing one's train and fear that it will rain at your garden party appear to lack the tendency to flee (or defend or protect oneself) and the bodily changes characteristic of this tendency, suggesting that they are not subserved by an affect program of fear or by any other affect program for that matter. By contrast, fear as manifested in a startle response after a loud bang has been invoked as an instance of emotion that bypasses appraisal (and therefore any appraisal-action tendency link). And the same has been suggested for euphoria from taking drugs.

To resolve the tension between fruitfulness and similarity, theorists have pursued different approaches. A first approach has been to prioritize *fruitfulness over similarity* by trimming the original, descriptive set of emotions, analogous to what happened to the definition of fish. Thus, some evolutionary theorists will argue that fear of missing one's train and fear that it will rain at your garden party simply do not qualify as emotions scientifically speaking (e.g., Griffiths & Scarantino, 2005). Other evolutionary theorists have opted to split the set of emotions into those that are and those that are not subserved by an affect program, similar to the set of jade being split into jadeite and nephrite (Griffiths, 2004). The most extreme form of prioritizing fruitfulness over similarity is undoubtedly the proposal to deny scientific status to the set of emotions altogether (as well as to its subsets). This is the option chosen by proactive theories such as the goal-directed theory as well as by some versions of psychological constructionism (i.e., Russell, 2003, 2009). This is analogous to the abolishment of air as a scientific set.

A second approach has been to prioritize *similarity over fruitfulness*. Theorists of various stripes have given up the ideal of finding an elegant set of necessary and sufficient conditions, and have turned instead to a cluster-type definition, also called a syndrome definition (Averill, 1980; Clore & Ortony, 2000; Parkinson, 1995; Scherer, 2005, 2022; Shaver et al., 1987). This is analogous to the definitions of lemons and many other categories such as species, mental disorders, and games. Theorists who have chosen this path still hold on to the explanation-based criteria for membership discussed in the previous sections, but they have relaxed the status of these criteria from necessary to typical. The criteria are most applicable to prototypical emotions and less to emotions on the fringe. Cluster-type definitions are convenient in that they allow for keeping the boundaries of the set of emotions fuzzy and malleable. Theorists who are weary of definitional discussions appear to embrace these cluster-type definitions as a welcome rescue from the definitional swamp. That this comes at a cost of fruitfulness may not strike them as particularly problematic because they tend to prioritize the study of prototypical emotions anyway – after all, one has to start somewhere.

Before closing, it should be acknowledged that few theorists distinguish between descriptive and scientific definitions, over even recognize the need to do so (but see e.g., Widen & Russell, 2010). Accordingly, the descriptions of their definitional approaches presented here should be understood as interpretations of the rationale behind their views. This contribution not only tries to underscore the importance of this distinction, but also

serves as an open invitation for the discussed theorists to make their positions more explicit or sharpen them, and to refine or ultimately overthrow the interpretations presented here.

Further Reading

Barrett, L. F., & Lida, T. (2024). Constructionist theories of emotions in psychology and neuroscience. In A. Scarantino (Ed.) *Emotion Theory: The Routledge comprehensive guide* (pp. 350-387). Routledge. https://doi.org/10.4324/9781315559940-20

Mesquita, B., & Parkinson, B. (2024). Social constructionist theories of emotion. In A. Scarantino (Ed.), *Emotion theory: The Routledge comprehensive guide* (pp. 388-407). Routledge. https://doi.org/10.4324/9781315559940-21

Moors, A. (2022). *Demystifying emotions: A typology of theories in psychology and philosophy*. Cambridge University Press. https://doi.org/10.1017/9781107588882

Moors, A. (2024). An overview of contemporary theories of emotions in psychology. In A. Scarantino (Ed.), *Emotion theory: The Routledge comprehensive guide* (pp. 232-259). Routledge. https://doi.org/10.4324/9781315559940-15

Moors, A., Ellsworth, P., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Scarantino, A. (2012). How to define emotions scientifically. *Emotion Review, 4*(4), 358–368. https://doi.org/10.1177/1754073912445810

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Shiota, M. N. (2024). Basic and discrete emotion theories. In A. Scarantino (Ed.), *Emotion theory: The Routledge comprehensive guide* (pp. 310-330). Routledge. https://doi.org/10.4324/9781315559940-18

Widen, S. C., & Russell, J. A. (2010). Descriptive and prescriptive definitions of emotion. *Emotion Review*, *2*(4), 377-378. https://doi.org/10.1177/1754073910374667

References

Arnold, M. B. (1960). *Emotion and personality*. Columbia University Press.

Armon-Jones, C. (1991). *Varieties of affect*. Harvester Wheatsheaf. https://doi.org/10.3138/9781487571887

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (Vol. 1, pp. 305–339). Academic Press. https://doi.org/10.1016/b978–0-12-558701-3.50018-1

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413–429. https://doi.org/10.1037/a0027555

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, *6*(4), 292–297. https://doi.org/10.1177/1754073914534479

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsx060

Barrett, L. F., & Lida, T. (2024). Constructionist theories of emotions in psychology and neuroscience. In A. Scarantino (Ed.), *Emotion theory: The Routledge comprehensive guide* (pp. 350–387). Routledge. https://doi.org/10.4324/9781315559940-20

Barrett, L. F., Lindquist, K. A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., & Brennan, L. (2007). Of mice and men: Natural kinds of emotions in the mammalian brain? A response to Panksepp and Izard. *Perspectives on Psychological Science*, *2*(3), 297–312. https://doi.org/10.1111/j.1745-6916.2007.00046.x

Boyd, R. N. (2010). Homeostasis, higher taxa, and monophyly. *Philosophy of Science*, *77*(5), 686–701. https://doi.org/10.1086/656551

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

Broekens, J., Jacobs, E., & Jonker, C. M. (2015). A reinforcement learning model of joy, distress, hope and fear. *Connection Science*, *27*(3), 215–233. https://doi.org/10.1080/09540091.2015.1031081

Buabang, E. K., Köster, M., Boddez, Y., Van Dessel, P., De Houwer, J., & Moors, A. (2023). A goal-directed account of action slips: The reliance on old contingencies. *Journal of*

Experimental Psychology: General, *152*(2), 496-508. https://doi.org/10.1037/xge0001280

Carnap, R (1947). *Meaning and necessity: A study in semantics and modal logic*. University of Chicago Press.

Clore, G. L., & Ortony, A. (2000). Cognition in emotion: Always, sometimes, or never? In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 24–61). Oxford University Press.

Colombetti, G. (2009). From affect programs to dynamical discrete emotions. *Philosophical Psychology*, *22*(4), 407–425. https://doi.org/10.1080/09515080903153600

Colombetti, G., & Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology*, *28*(8), 1157–1176. https://doi.org/10.1080/09515089.2014.976334

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Wiley. https://doi.org/10.1002/0470013494.ch3

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. Academic Press. https://doi.org/10.1016/c2009–0-02293-4

Fridlund, A. J. (2017). The behavioral ecology view of facial displays, 25 years later. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 77–92). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190613501.003.0005

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H., & Parrott, W. G. (2011). Basic emotions or ur-emotions? *Emotion Review*, *3*(4), 406–415. https://doi.org/10.1177/1754073911410742

Griffiths, P. E. (2004). Emotions as natural and normative kinds. *Philosophy of Science*, *71*(5), 901–911. https://doi.org/10.1086/425944

Griffiths, P. E., & Scarantino, A. (2005). Emotions in the wild: The situated perspective on emotion. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 347–353). Cambridge University Press. https://doi.org/10.1017/cbo9780511816826.023

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*, 1–26.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge University Press. https://doi.org/10.1017/cbo9781139174138

Ierodiakonou, K. (1993). The Stoic division of philosophy. *Phronesis*, *38*(1), 57–74. https://doi.org/10.1163/156852893321052451

James, W. (1890). *The principles of psychology*. Vol. 2. Holt.

Keltner, D., & Haidt, J. (2001). Social functions of emotions. In T. J. Mayne & G. A. Bonanno (Eds.), *Emotions: Current issues and future directions* (pp. 192–213). Guilford. https://doi.org/10.1080/026999399379168

Keltner, D., Haidt, J., & Shiota, M. N. (2006). Social functionalism and the evolution of emotions. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 115–142). Psychology Press. https://doi.org/10.4324/978020378296

Keltner, D., Tracy, J. L., Sauter, D., & Cowen, A. (2019). What basic emotion theory really says for the twenty-first century study of emotion. *Journal of Nonverbal Behavior*, *43*, 195–201. https://doi.org/10.1007/s10919–019-00298-y

Kreibig, S. D. (2010). Autonomic nervous activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Lang, P. J. (1994). The motivational organization of emotion: Affect-reflex connections. In S. H. M. van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 61–93). Erlbaum. https://doi.org/10.4324/9781315806914

Lange, J., Dalege, J., Borsboom, D., van Kleef, G. A., & Fischer, A. H. (2020). Toward an integrative psychometric model of emotions. *Perspectives on Psychological Science*, *15*(2), 444–468. https://doi.org/10.1177/1745691619895057

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

LeDoux, J. E. (2012b). Rethinking the emotional brain. *Neuron*, *73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

León, F., Szanto, T., & Zahavi, D. (2019). Emotional sharing and the extended mind. *Synthese*, *196*(12), 4847–4867. https://doi.org/10.1007/s11229–017-1351-x

Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review*, *6*(2), 100–112. https://doi.org/10.1177/1754073913512003

Leventhal, H. (1984). A perceptual-motor theory of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 117–182). Academic Press. https://doi.org/10.1016/s0065–2601(08)60119-7

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences*, *28*(2), 169–245. https://doi.org/10.1017/s0140525x0500004x

Lewis, M. D., & Liu, Z. X. (2011). Three time scales of neural self-organization underlying basic and nonbasic emotions. *Emotion Review*, *3*(4), 416–423. https://doi.org/10.1177/1754073911410748

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–143. https://doi.org/10.1017/s0140525x11000446

Longworth, F., & Scarantino, A. (2010). The disjunctive theory of art: The cluster account reformulated. *British Journal of Aesthetics*, *50*(2), 151–167. https://doi.org/10.1093/aesthj/ayq001

Manstead, A. S., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 221–232). Oxford University Press.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. Psychological Review, 98(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Mauss, I. B., Zerwas, F. K., Wilhelm, F. H., & John, O. P. (2024). Coherence of emotional response systems: Theory, measurement, and benefits. *Advances in Experimental Social Psychology*, 69, 59-149.

Mesquita, B. (2022). *Between us: How cultures create emotions*. Norton.

Mesquita, B., & Ellsworth, P. C. (2001). The role of culture in appraisal. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 233–248). Oxford University Press.

Moors, A. (2017a). Integration of two skeptical emotion theories: Dimensional appraisal theory and Russell's psychological construction theory. *Psychological Inquiry*, *28*(1), 1–19. https://doi.org/10.1080/1047840X.2017.1235900

Moors, A. (2017b). The integrated theory of emotional behavior follows a radically goal-directed approach. *Psychological Inquiry*, *28*(1), 68–75. https://doi.org/10.1080/1047840X.2017.1275207

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. https://doi.org/10.1177/1754073916669595

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331. https://doi.org/10.1037/0033-295X.97.3.315

Panksepp, J. (2012). In defense of multiple core affects. In P. Zachar & R. D. Ellis (Eds.), Categorical versus dimensional models of affect: A seminar on the theories of Panksepp and Russell (Vol. 7, pp. 31–78). John Benjamins. https://doi.org/10.1075/ceb.7.02pan

Parkinson, B. (1995). *Ideas and realities of emotions*. Routledge. https://doi.org/10.4324/9780203135648

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review*, *9*(4), 278–311. https://doi.org/10.1207/s15327957pspr0904_1

Parkinson, B. (2007). Getting from situations to emotions: Appraisal and other routes. *Emotion*, *7*(1), 21–25. https://doi.org/10.1037/1528-3542.7.1.21q

Parkinson, B. (2012). Piecing together emotion: Sites and time-scales for social construction. *Emotion Review*, *4*(3), 291–298. https://doi.org/10.1177/1754073912439764

Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Reisenzein, R. (2012). What is an emotion in the belief-desire theory of emotion? In F. Paglieri, L. Tummolini, R. Falcone, & M. Miceli (Eds.), *The goals of cognition: Essays in honor of Cristiano Castelfranchi* (pp. 181–211). College Publications.

Rolls, E. T. (2005). *Emotion explained*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570035.001.0001

Roseman, I. J. (2013). Appraisal in the emotion system: Coherence in strategies for coping. *Emotion Review*, *5*(2), 141–149. https://doi.org/10.1177/1754073912469591

Russell, J. A. (2009). Emotion, core affect, and psychological construction. *Cognition and Emotion*, *23*(7), 1259–1283. https://doi.org/10.1080/02699930902809375

Russell, J. A. (2012). From a psychological constructionist perspective. In P. Zachar & R. D. Ellis (Eds.), *Categorical versus dimensional models of affect: A seminar on the theories of Panksepp and Russell* (pp. 79–118). John Benjamins. https://doi.org/10.1075/ceb.7.03rus

Russell, J. A. (2014). My psychological constructionist perspective, with a focus on conscious affective experience. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 183–208). Guilford.

Sauter, D. A., & Russell, J. A. (2025). What do nonverbal expressions tell us about emotion? In A. Scarantino (Ed.), *Emotion theory: The Routledge comprehensive guide* (pp. 543–560). Routledge.

Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1, pp. 49–80). Academic Press. https://doi.org/10.1016/S0065–2601(08)60048-9

Scherer, K. R. (1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89–126). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-2792-6_4

Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M. D. Lewis & I. Granic (Eds.), *Emotion, development and selforganization: Dynamic systems approaches to emotional development* (pp. 70–99). Cambridge University Press. https://doi.org/10.1017/cbo9780511527883.005

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Scherer, K. R. (2009). Emotions are emergent processes: They require a dynamic computational architecture. *Philosophical Transactions of the Royal Society: B, 364*, 3459–3474. https://doi.org/10.1098/rstb.2009.0141

Scherer, K. R. (2022). Theory convergence in emotion science is timely and realistic. *Cognition and Emotion, 36*(2), 154–170. https://doi.org/10.1080/02699931.2021.1973378

Shand, A. F. (1914). *The foundations of character: Being a study of the tendencies of the emotions and sentiments*. Macmillan. https://doi.org/10.1037/10666-000

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061

Searle, J. R. (1958). Proper names. *Mind, 67*(266), 166–173. https://doi.org/10.1093/mind/lxvii.266.166

Shariff, A. F., & Tracy, J. L. (2011b). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395–399. https://doi.org/10.1177/0963721411424739

Solomon, R. C. (1993). *The passions: Emotions and the meaning of life*. Hacket.

Shiota, M. N. (2024). Basic and discrete emotion theories. In A. Scarantino (Ed.) *Emotion Theory: The Routledge comprehensive guide* (pp. 310-330). Routledge. https://doi.org/10.4324/9781315559940-18

Swartz, N. (1997). *Definitions, dictionaries, and meanings.* www.sfu.ca/~swartz/definitions.htm

Tappolet, C. (2016). *Emotions, values, and agency*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199696512.001.0001

Watson, J. B. (1919). A schematic outline of the emotions. *Psychological Review*, *26*(3), 165–196. https://doi.org/10.1037/h0072509

Wittgenstein, L. (1953). *Philosophical investigations*. Macmillan.

Wood, A. & Coan, J.A. (2023). Beyond nature versus nurture: The emergence of emotion. *Affective Science, 4*, 443–452. https://doi.org/10.1007/s42761-023-00212-2

Notes

[i] Note that elimination may also come in the form of the subsumption of the set into another, overarching, set. An example is the subsumption of individual forces (e.g., gravity) under the broader set of energy (i.e., the capacity to do work).

[ii] Building on the Stoic tradition of systematic division as discussed by Ierodiakonou (1993), one might interpret a divisio definition as a type of extensional definition where the extension of a set is not a list of individual members, but a list of subsets.

[iii] An exception is Panksepp's (2012) selection. His set of basic emotions (i.e., fear, rage, panic/grief, lust, care, play, and seeking) was not based on language but rather on available evidence for neural circuits in animals. Here too, however, some arbitrariness may have crept in (e.g., who decides which circuits to study?).

[iv] It is fair to note that some appraisal theorists, such as Scherer (1988) do not think goal relevance is strictly necessary but allow for cases in which molecular appraisals of intrinsic valence and novelty are sufficient. For Scherer (2000), moreover, the mark of the emotional is a high degree of synchronization among components. For Frijda (1986), the mark of the emotional is control precedence. Moors (2022) has argued that both a high degree of synchronization and control precedence (which she considers to be explananda) can be explained by a high degree of goal relevance (which she considers to be the explanans).

[v] Note that Parkinson (2012) counts as a social constructionist theorist because he subscribes to the general idea that emotions are constructions in the sense that they are made up of mechanisms that are not emotional per se. These mechanisms may not be restricted to the categorization of bodily feelings as proposed by Barrett (2006) but may also include distributed appraisal processes.

[vi] For an extensive discussion of similarity, that is, how well each of the theories can account for each of the desiderata from the descriptive definition, the reader is referred to Moors (2022).