Check for updates

emotion review

Emotion Review
Vol. 18, No. 2 (April 2026) 61–75
© The Author(s) 2026
ISSN 1754-0739
DOI: 10.1177/17540739261422553
https://journals.sagepub.com/home/emr

Emotions as High-Impact Decisions: A Goal-Directed Theory

Agnes Moors
Research Group of Quantitative Psychology and Individual Differences, Center for Social and Cultural Psychology, KU Leuven—University of Leuven, Leuven, Belgium

Abstract

I defend the adequacy of the goal-directed theory (GDT) of emotion. After outlining the properties of emotions (mental and bodily aspects, valence, intensity, urgency, automaticity, and irrationality) and common emotion types that emotion theories seek to explain, I contrast the GDT with traditional emotion theories, portraying the former as proactive and the latter as reactive. The GDT proposes that most behavior and feelings, emotional or otherwise, arise from multiple competing, hierarchical goal-directed cycles, each involving discrepancy detection and strategy/behavior selection. I systematically assess the GDT's ability to account for these properties and emotion types. Finally, I discuss implications for (a) the scientific status of emotions, (b) their influence on other entities, (c) emotion regulation, and (d) agency and moral responsibility.

Keywords

emotion, goal-directed, instrumental, action, irrational

This article describes and defends the adequacy of the goal-directed theory (GDT) of emotion, developed across a series of papers (Moors, 2017a, 2017b, 2023, 2024, 2025; Moors et al., 2017, 2021; Moors & Boddez, 2017; Moors & Fischer, 2019) and book (Moors, 2022). The theory is a mechanistic sketch: It takes a clear position about the mechanisms involved in the phenomena we tend to call emotions in terms of the types of information that are processed, while being more liberal in terms of the format in which this information is represented and the operations at play. Also, various lower-level mechanistic details remain to be worked out, as do the neuroscientific underpinnings. The further development of the theory in these regards will have to be done under the impulse of empirical research.

The theory stems from a dissatisfaction with traditional emotion theories and presents itself as an alternative to these theories. In order to arrive at a clear delineation between the GDT and these traditional theories, let me take a step back and clarify what it is that emotion theories try to accomplish. Emotion theories take emotions as the phenomenon to be explained (i.e., explanandum), and propose explanations for them (i.e., linking the explanandum to an explanans). Ideally, the start of this endeavor is a working definition of emotions, which helps us draw the provisional contours of the set of entities we call emotions. Working definitions come in the form of typical and apparent properties of emotions and sometimes in the form of emotion types. A working definition specifies what it is about emotions that we want to see explained—the so-called desiderata of emotion theories. Next, theorists develop explanations. These explanations not only try to increase our understanding of the things we call emotions, but also seek to provide a deep common ground that is shared by all members of the set. In other words, explanations seek to generate criteria that allow demarcating the set of emotions from other sets in a more principled manner than just through typical and apparent properties. If an explanation is successful in this regard, it can elevate the set of emotions to a scientific kind. Of course, the adequacy of an explanation not only depends on its capacity for demarcation but also on whether the explanation itself is valid, that is, whether the explanans

Corresponding author: Agnes Moors, Research Group of Quantitative Psychology and Individual Differences, Center for Social and Cultural Psychology, KU Leuven—University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium.
Email: agnes.moors@kuleuven.be

exists and can be linked to the explanandum. This will often require empirical validation.

In the following sections, I start from a working definition of emotions, listing a set of typical and apparent properties as well as typical emotion types for which there is fair consensus in the emotion-scientific community. I then turn to the constitutive and causal-mechanistic explanations proposed by theories in this community. While constitutive explanations can account for some desiderata, causal-mechanistic explanations are required for others. I clarify how the causal-mechanistic explanations of traditional emotion theories—characterized as reactive theories—are able to meet these other desiderata. Next, I clarify how the causal-mechanistic explanation of the GDT—characterized as a proactive theory—can meet them as well. Although traditional theories and the GDT are nearly equivalent in this regard, they diverge in the implications they carry for the scientific status of emotions. The mechanisms of traditional theories are special-purpose mechanisms that can ground emotions as a distinct scientific set, provided that they exist. Those of the GDT are general-purpose mechanisms that are shared between emotional and non-emotional phenomena and therefore cannot confer scientific status on the set of emotions. Other implications discussed are how the lack of scientific status impacts research on the influence of emotions on other entities, and how the GDT may better equip individuals to regulate and take responsibility for their "emotions" than traditional theories.

Working Definition

The typical and apparent properties of emotions for which there is fair consensus are the following. First, emotions have mental and bodily aspects. The mental aspect refers to the fact that emotions have an object or are about something, a property called object-directedness or Intentionality.[1] One can be angry about something or afraid of something. An organism can be directed at something by forming a mental representation of it. Thus, mental activity is often understood as representation-mediated processing. While much of this mental activity occurs below the surface of awareness, part of it will be consciously experienced or felt. This is called the experiential aspect of emotions. The bodily aspect refers to the fact that it often involves physiological responses as well as subtle (i.e., expressive) and/or coarse (i.e., musculoskeletal) behavior. If overt behavior is absent, emotions can still involve an action tendency or readiness to act, which qualifies as a mental representation with behavioral content. One can have the tendency to fight or flee, for instance.

Second, emotions do not involve just any type of mental and bodily activity, but activity with a certain "heat." The heat of emotions refers to their positive or negative quality and intense quantity. Third, emotions are often triggered automatically, in the sense that they arise quickly, unintentionally, and efficiently, and/or are difficult to counteract (Moors, 2016). A fourth property of emotions is their urgency or control precedence, meaning that they take priority over mundane goal pursuit. A fifth property of emotions is that they have an irrational flavor insofar as they can bias our thoughts and lead to maladaptive behavior.

Moving to the emotion types in need of an explanation, consensus seems to be strongest for anger, fear, sadness, and joy, along with their shades (e.g., rage, panic, despair, satisfaction, and euphoria) and slightly weaker for surprise, disgust, pride, shame, guilt, envy, jealousy, awe, interest, wonder, and so on.

Explanations

Two types of explanations are commonly offered for the emotions in the working set: constitutive explanations, which specify the components of emotions, and causal-mechanistic explanations, which specify causal factors and mechanisms. The constitutive explanations of emotion theories range from being mono- to multi-componential. Guided by the desire to be parsimonious, some theories try to include as few components as possible inside the emotion. Some theories identify emotions with feelings, others with cognitions, and still others with action tendencies. At the other end of the spectrum are theories that conceptualize emotions as episodes with many components, including feelings, cognitions, action tendencies, physiological responses, and expressive behavior. Several of these components can fulfill some of the desiderata specified earlier: The feeling component delivers the experiential aspect, the cognitive component the Intentional aspect, and the motivational, somatic, and motor components the embodied aspects. But what about the other desiderata?

Some theories (often philosophical ones) account for the remaining desiderata, such as the heat of emotions, by stating that the component(/s) that they identify with emotions is(/are) somehow special. For instance, feeling theories propose that emotions are special types of feelings, cognitive theories that they are special types of cognitions, and motivational theories that they are special types of action tendencies. The solution offered by other theories (often psychological ones) is to state that the components of their constitutive explanation are generated by a special type of mechanism or information process.[2]

Focusing on theories offering causal-mechanistic explanations, I have proposed to organize them into two broad classes: reactive theories, offering a stimulus-driven mechanism, and proactive theories, offering a goal-directed mechanism. Many traditional theories are reactive. My own GDT is one instance of a proactive theory. Let us briefly state the essence of reactive theories before diving into the GDT.

Reactive Theories

Reactive theories build on the mechanism for emotion causation proposed by evolutionary or basic emotions theories (e.g., Ekman & Cordaro, 2011; Keltner et al., 2019; Shiota,

2024; Tracy & Randles, 2011; see Moors, 2022, Chapter 4). This mechanism takes the form of a stimulus-driven process. Incoming stimuli activate an innate association between a stimulus representation ([S]) and a response representation or action tendency ([R]). Such [S-R] associations, also known as affect programs, are supposed to be types of instincts, that is, remnants of the behavioral solutions that our evolutionary ancestors selected to deal with recurrent challenges and opportunities in their environment in the past. The associations have been framed by some theories in concrete terms, as in the example "snake-flee," and by others in more abstract terms, as in the example "danger-protection." In the concrete variants, a snake is perceived in concrete terms—as a snake—and this in turn activates a fixed link between the representation of the snake and the concrete tendency to flee. Once the affect program is activated, it subsequently triggers physiological responses, preparing the body for fleeing, characteristic facial and vocal expressions, actual fleeing behavior, and feelings of fear. In the abstract variants, the snake is perceived in abstract terms—as dangerous—before activating a fixed link between the representation of danger and the abstract tendency to protect oneself. In the latter variants, the affect program is preceded by a stimulus evaluation process, also known as an appraisal process (see also Box 1).

Reactive theories are dual-process theories in that they contrast the stimulus-driven process responsible for emotional behavior in emotional episodes with a goal-directed process for instrumental behavior in non-emotional episodes. In a goal-directed process, the available behavior options are compared ([$S:R_1\text{-}O^v$, $S:R_2\text{-}O^v$]) and the option with the highest expected utility is selected. The expected utility of a single behavior option equals the value of the outcomes (i.e., values) of the behavior, weighted by the subjective probabilities (i.e., expectancies) that these outcomes will indeed occur. To illustrate, consider the contrast often drawn between emotional and instrumental aggression. A father may yell at his child out of anger (i.e., emotional aggression) or as a strategy to make the child obey (i.e., instrumental aggression). Reactive theories assume that emotional aggression is elicited by a stimulus-driven process: The stimulus activates the innate affect program for anger, for instance, an "offense-aggression" link, which then gets implemented in overt aggression. They contrast this with instrumental aggression, in which the father selects aggressive behavior because he expects it to be the most effective means for making the child obey.

While reactive theories hold that stimulus-driven processes account for the pure emotion, they still reserve a role for goal-directed processes in planning and emotion regulation. Planning refers to the translation of a more abstract action tendency into a more concrete one. If the father has the tendency to be aggressive, he still has to select a concrete action tendency such as yelling, hitting, or throwing daggers. And if the father has the tendency to yell, he may also want to select a particular loudness and content. Emotion regulation refers to the (more or less successful) attempt to modify or maintain a current or anticipated emotion. If the father experiences anger or a tendency to aggress, which goes against his goal to maintain a good relationship with the child, he may choose to regulate his anger, for instance, by suppressing his aggressive urge or leaving the room.

Importantly, reactive theories characterize the interplay between stimulus-driven and goal-directed processes as *default-interventionist.* The stimulus-driven process involved in the pure emotion is the default, whereas the goal-directed process involved in planning and/or regulation can only supplement and/or intervene in this process. This division of labor is rooted in the deep-seated conviction that there is a trade-off between automaticity and adaptiveness (akin to the trade-off between speed and accuracy), which is tied to the complexity of processes. The story goes that stimulus-driven processes are simple and thus operate automatically—that is, when operating conditions are poor (i.e., a lack of time, attention, and motivation). This is why they operate by default and are difficult to counteract. On the other hand, the simple nature of stimulus-driven processes also entails that they are rigid ([S-R] links cannot adapt to changes in behavioral outcomes), and are therefore more likely to produce maladaptive behavior. Goal-directed processes, on the other hand, are more complex and therefore also more flexible ([$S:R\text{-}O^v$] representations can adapt to changes in behavioral outcomes). This is why they are well equipped to produce adaptive behavior, enabling them to intervene when emotional action tendencies generated by stimulus-driven processes go astray. On the other hand, the more complex nature of goal-directed processes also entails that they are nonautomatic, which is why they can only play their intervening role when operating conditions are ample (i.e., abundant time, attention, and motivation).

Adding this all up, reactive theories can make perfect sense of the desiderata that emotions are triggered automatically in the sense of being fast, efficient, unintentional, and/or difficult to counteract. In addition, these theories account for the irrational flavor of emotions, particularly their tendency to lead to maladaptive behavior. Reactive theories add nuance to this idea, however, suggesting that emotions can still be adaptive if the action tendency activated as part of the ancestral [S-R] link proves instrumental to the person's current goals. This instrumentality depends on whether the outcome structure of the current environment resembles that of the ancestral environment where the [S-R] link was first formed. Thus, fleeing from a snake without forethought of the outcomes of this behavior is adaptive in the wild, where it may save a person's life, but not in the zoo, where it may lead to being ridiculed by others.

Reactive theories not only account for the typical properties ascribed to emotions, but also for the emotion types

populating the working set of emotions. Evolutionary theories, for instance, postulate that there is a unique affect program for a limited number of so-called basic emotions (e.g., fear, anger, joy, sadness, surprise, and disgust).[3] Each affect program corresponds to an innate [S-R] link where the S-part reflects a challenge or opportunity and the R-part reflects a time-tested solution of our ancestors that became hard-wired into the brain of our species.

Given that affect programs are dedicated to emotions, they are in principle capable of buying the set of emotions scientific status. An episode counts as emotional if it is undergirded by an affect program; everything else is not an emotion. Such a scientific definition may be partly revisionary if it keeps only some original members of the working set of emotions on board (e.g., full-blown fear and anger) while throwing out others (e.g., when someone says they "fear that they disagree"). But the real problems are elsewhere. The first problem is that the boundary between emotional instincts known as affect programs (e.g., danger-flight and offense-fight) and non-emotional instincts (e.g., hunger-foraging and infant-nurturing) remains artificial. Thus, while emotional instincts have been invoked to provide a biological basis to the set of emotions, we ultimately have to rely on language to determine which instincts qualify as emotional and which do not, and by inference, where to draw the boundaries of the set of emotions.[4]

The second problem is empirical. Affect programs can only serve as a common denominator of the set of emotions if they exist. Empirical efforts to obtain direct and indirect evidence for the existence of affect programs are abundant and ongoing (see reviews by Ekman, 1992; Keltner et al., 2019; Kreibig, 2010; Levenson, 2014; Shiota, 2024) but according to opponents, not convincing and open to alternative interpretations (Barrett, 2006; Lindquist et al., 2012; Ortony & Turner, 1990; Russell, 2003; see also Moors, 2022, Chapter 4).

Proactive Theories

Proactive or instrumental theories of emotion comprise reinforcement learning theories of emotions (Rolls, 2005; see also Broekens et al., 2015) as well as my own GDT (Moors 2017a, 2017b; Moors et al., 2017), which is the focus here. The GDT qualifies as a dual-process theory in that it acknowledges the existence of both stimulus-driven and goal-directed processes, yet it diverges from reactive theories in its conception of the interplay between these processes and their respective explanatory territories.

The GDT characterizes the interplay between stimulus-driven and goal-directed processes as *parallel-competitive*. In sharp opposition to traditional dual-process theories, it assumes that not only stimulus-driven but also goal-directed processes can be automatic. This means that both of them can operate under poor operating conditions (i.e., when time, attention, and motivation are scarce), which entails that they will often operate in parallel and thereby compete with each other. In addition, the theory postulates that goal-directed processes are stronger than most stimulus-driven ones and will therefore often win the competition. The net result is that most of our behavior will be caused by goal-directed processes.

The GDT proposes that goal-directed processes are not merely involved in instrumental behavior or emotion regulation, but are also at the heart of emotions themselves. To advance this argument, I will proceed as follows. In the upcoming section, I will lay out the steps in the goal-directed process proposed by the GDT, and I will illustrate these steps using mundane, "non-emotional" examples. To bring out the analogy between non-emotional and emotional episodes, I will proceed to a reanalysis of the contrast between instrumental and emotional aggression through the lens of the GDT. This will set the stage in the next sections for a systematic examination of how well the GDT is able to meet the desiderata for adequacy specified earlier, both in terms of properties and emotion types. To assist in this exercise, corollary assumptions of the GDT will be introduced along the way.

Nuts and Bolts of the GDT

The theory goes beyond earlier proactive theories in that it considers the goal-directed process involved in behavior selection as one stage within a broader goal-directed cycle, where it is preceded by a stage of discrepancy detection (see Figure 1, Panel a). Thus, the cycle starts with the comparison between the representation of a stimulus ([S]) and a first goal (representation of a valued outcome; [O^v]). If the person detects a discrepancy, a second goal is activated aimed at reducing the discrepancy. The second goal can be achieved via three broad strategies: The person can (a) choose a behavior from their behavior repertoire (i.e., assimilation), (b) swap the first goal by a different one (i.e., accommodation), or (c) reinterpret the stimulus so that it appears less discrepant with the first goal (i.e., immunization). The GDT assumes that the selection of the broad strategies of assimilation, accommodation, and immunization, as well as the selection of specific behavior options in the case of assimilation, are determined by the weighing up of the expected utilities of these respective strategies and behavior options. In the case of assimilation, the selected behavior option activates its corresponding action tendency, which is a third goal (i.e., a behavioral goal: [R]). This tendency is implemented in overt behavior (R), resulting in an actual outcome (O). The actual outcome is then fed back as the input to the next run of the cycle. The cycle is repeated until the discrepancy is effectively reduced or until a discrepancy with a more valued goal is detected and clamors priority.

To illustrate, imagine a cook who wants to obtain the perfect bearnaise (first goal; [O^v]) but notices that the eggs

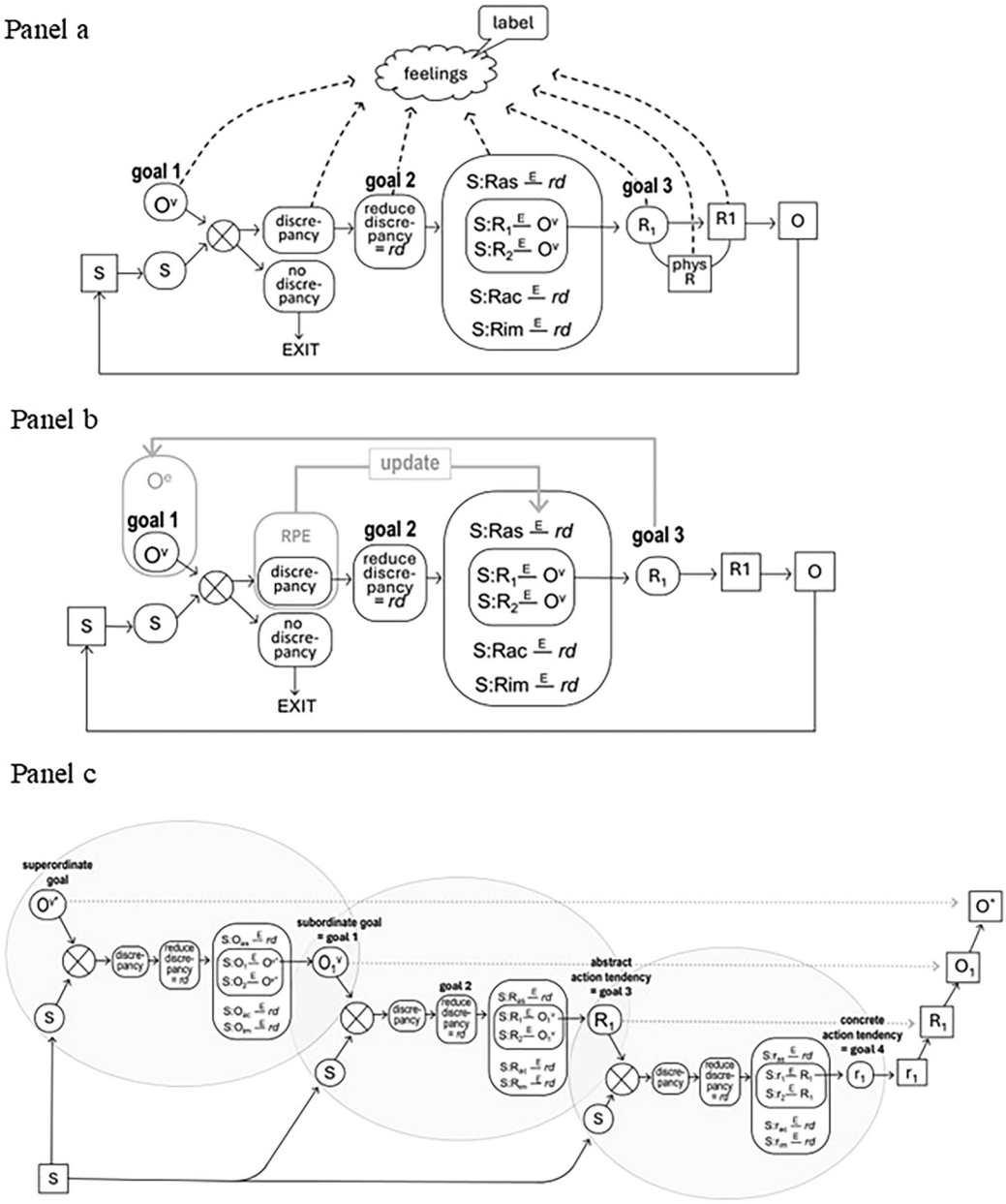

Figure 1. Goal-directed cycle.

Note. Adapted from Moors, Demystifying emotions. Copyright (2025), with permission from *Cambridge University Press*.

have started to set (stimulus; [S]). To reduce this discrepancy, she can choose to engage in behavior (i.e., assimilation), in which case she still has to decide whether she will try to salvage the sauce by making it more liquid (behavior option 1; [R_1]) or restart the whole process from scratch (behavior option 2; [R_2]). She can also decide to change the menu and serve yesterday's tomato sauce instead (i.e., accommodation). Finally, she can re-evaluate the consistency of the sauce as liquid enough and pretend all is going according to plan (i.e., immunization).

The GDT introduces at least five nuances to this basic description. First, the stimulus representation ([S]) at the start of the cycle need not be a current state but can also be an anticipated state. For instance, the presence of dark clouds may cause a sunbather to anticipate that it will rain soon. The rain may constitute an anticipated discrepancy with his first goal to get a tan ([O^v]). The anticipated discrepancy in turn activates a second goal to reduce the discrepancy. Note that the second goal can be called a protection or prevention goal, in contrast with the progress or promotion goal following a current discrepancy. Reducing the anticipated discrepancy can be done by selecting between the available behavior options (i.e., assimilation) or through accommodation or immunization. The sunbather may go to the nearest sun palace ([R_1]), or he may use artificial bronzer ([R_2]). He may also give up his goal to get a tan for the day (i.e., accommodation), or he may reinterpret his current tan as deeper than it really is (i.e., immunization).

Second, the representation that gets compared to the first goal need not even be a stimulus or *external* event, but can also be an *internal* event, such as the representation of negative affect. If this affect is discrepant with the goal to have positive affect, emotion regulation may be selected as a form of assimilation (see Moors, 2022, p. 238).

Third, the GDT follows the principles for reinforcement learning to update expectancies of behavior options (see Figure 1, Panel b; Moors, 2022). When the action tendency in a first run of the cycle is activated, it not only translates into overt behavior producing an actual outcome, but also generates an expectation that the valued outcome will in fact occur. This entails that in the next run of the cycle, the actual outcome gets compared to an outcome that is not only valued but also expected. Any discrepancy between the two elements qualifies as what is known as a "reward prediction error." This reward prediction error can be used to update the expectancy of the formerly selected behavior option in the action repertoire. If the cook's attempt to salvage the bearnaise ([R_1]) does not result in the valued and expected outcome, the expectancy of this behavior for this outcome will be reduced, so that she may now choose to start from scratch ([R_2]).

Fourth, the GDT embraces two further sources of complexity. Horizontal complexity refers to the fact that people have multiple, sometimes competing goals. This entails that the expected utility of each behavior option must not only be calculated for the first goal in the cycle, but also for other goals that are considered. The cook's choice between salvaging the present sauce ([R_1]) or starting anew ([R_2]) will not only depend on how effective these behaviors are in producing the perfect bearnaise ([O^v_1]) but also on how effective they are in satisfying other goals, such as the goal to serve the customers on time ([O^v_2]) and the goal to preserve effort ([O^v_3]). In other words, action selection may depend on a more extensive cost–benefit analysis.

Vertical complexity refers to the idea that goal-directed cycles are organized in a hierarchy (see Figure 1, Panel c). This means that the first goal in a cycle ([O^v]) was itself selected because it had the highest expected utility for a higher-order goal ([O^{v*}]). So the goal to obtain the perfect bearnaise ([O^v]) may itself have been chosen because the cook expected it to have a high expectancy of leading to the valued outcome of receiving a Michelin star ([O^{v*}]). Note that the hierarchy not only extends upwards but also downwards. Once the third goal of salvaging the sauce ([R]) has been selected, this goal is not yet fulfilled (i.e., it is still discrepant with the current state) and the cook still has to choose between whisking in a splash of cold water ([r_1]) or a dash of melted butter ([r_2]).

To reiterate, the theory assumes that the goal-directed cycle involved in episodes of mundane goal pursuit is the same as that involved in emotional episodes. To get this point across, let us return to the cases of instrumental and emotional aggression and reanalyze them using the goal-directed cycle. In the instrumental case, the cycle starts with the child's unruly behavior ([S]), which constitutes a discrepancy with the father's goal of instilling obedience in the child ([O^v]). To reduce this discrepancy, the father chooses to engage in yelling ([R_1]) instead of begging ([R_2]) or instead of giving up his goal for obedience (i.e., accommodation) or reinterpreting the child's behavior as a joke (i.e., immunization).

In the emotional case, the cycle also starts with the discrepancy between the child's unruly behavior and one of the father's goals, for instance, his goal to exert power ([O^v]). This aligns with what traditional theories would call an appraisal of the child's behavior as an insult or offense. According to these traditional theories, however, the next step is that the appraisal of insult or offense directly triggers the tendency to behave aggressively. The GDT instead proposes that the father *chooses* to reduce the discrepancy by engaging in the aggressive behavior of yelling ([R_1]) because he evaluates this behavior as having the highest expectancy for regaining power, and in fact, as having the highest total expected utility (i.e., the most favorable cost–benefit ratio).

If emotional and non-emotional episodes indeed share the same mechanism, how can the theory still account for the desiderata, that is, the typical and apparent properties ascribed to emotions in the working definition? This is discussed in some depth in the next section.

Box 1. The difference between the GDT and appraisal theories of emotion.

I have argued that appraisal theories as we know them today qualify as *stimulus evaluation* theories, whereas the GDT counts as a *response evaluation* theory. To be more precise, the GDT combines rudimentary stimulus evaluation—in the discrepancy detection stage—with response evaluation—in the strategy/behavior selection stage. Moreover, this rudimentary stimulus evaluation in the GDT aligns with so-called primary appraisal in appraisal theories (i.e., the presence/absence of a discrepancy in the GDT aligns with the appraisals of goal in/congruence, and the magnitude of the discrepancy and goal value contribute to goal relevance in appraisal theories). While it may be tempting to assume that the response evaluation in the GDT also aligns with so-called secondary appraisal in appraisal theories, they are not identical. Response evaluation in the GDT assesses the expected utilities of specific strategies/behavior options (i.e., a mechanism with "strategy/behavior option" as input and "expected utility" as output). For instance, it assesses whether fleeing has a higher expected utility than fighting. Secondary appraisal in appraisal theories, on the other hand, assesses the degree of controllability of the stimulus (i.e., a mechanism with "stimulus" as input and "degree of controllability" as output). For instance, it assesses whether the barking dog is easy or difficult to control. It might be conjectured that the controllability appraisal ultimately relies on the expected utilities of specific behavior options, and hence that appraisal theories do not substantially differ from the GDT. Such a conclusion, however, would overlook another important difference: Appraisal theories hold that the pattern of appraisals, including low/high controllability, fixes the initial action tendency that is elicited. For instance, if the dog (which is goal-incongruent) is easy/difficult to control, the tendency to fight/flee will be elicited (characteristic of anger/fear). In the GDT, on the other hand, the specific behavior option with the highest expected utility activates its corresponding action tendency, without making a detour via an appraisal of the degree of controllability in terms of easy vs. difficult. This difference between the two theories may lead to different predictions (see Moors, 2022; Moors & Fischer, 2019), which can be tested in empirical research. To put it simply, if a person has *high control* in the sense that they can *flee* (fleeing has the highest expected utility), appraisal theory would predict an initial tendency to *fight* whereas the GDT would predict an initial tendency to *flee* (see Fischer et al., 2020a, 2020b; Moors et al., 2019; see also Moors, 2017a, p. 10, for an in-depth discussion).

Another reason for failing to distinguish the GDT from appraisal theories is the notion of action tendencies. Some authors seem to assume that when a theory confers a central place to action tendencies in emotional episodes, it counts as goal-directed (e.g., Eder, 2023). Guarding against the overinclusive use of the term goal-directed, however, I have argued that an action tendency, although defined as a behavioral goal, does not count as a goal-directed process (Moors, 2023). Such a process involves the computation or activation of one or more behavior options for one or more valued outcomes (Heyes & Dickinson, 1990). Critically, these valued outcomes should be external to the outcomes that define the behavior itself. Thus, saying that fleeing is caused by the tendency/goal to flee is not enough because the goal to flee defines the behavior of fleeing. For fleeing to be considered as caused by a goal-directed process, the tendency to flee must be caused by a process in which the expectancy of fleeing for a different goal—such as safety—was estimated as high.

The GDT and Emotional Properties

After discussing the general properties of emotions that they are characterized by mental and bodily changes, I move on to the properties that differentiate emotions from other mental and bodily phenomena—valence, intensity, automaticity, urgency, and irrationality—and I end with illustrations of how the GDT can make sense of emotion types from natural language.

Mental and Bodily Aspect

Traditional theories account for the mental and bodily aspects of emotions by treating them as components of emotional episodes. Each of these components has an equivalent in the goal-directed cycle of the GDT: The stages of discrepancy detection and action selection correspond to the cognitive component and secure the Intentionality of emotions. The result of the action selection stage—the action tendency—maps onto the motivational component, the ensuing overt behavior maps onto the motor component, and the physiological activity that prepares for and supports this overt behavior onto the somatic component. Together, the motivational, motor, and somatic components secure the bodily aspect of emotions.

Even though all the steps in the goal-directed cycle can operate automatically and below the surface of awareness, they require a certain level of activation to operate, and some steps can become conscious. Mental representations (rounded boxes in Figure 1) must exceed a first threshold of activation to propel their activation to the next step (i.e.,

be causally efficacious) and a second threshold of activation for their content to become conscious (Moors, 2016). Non-mental entities (i.e., overt behavior, somatic responses, and actual stimuli or outcomes; rectangular boxes in Figure 1) have to be represented before they can become conscious (Moors, 2017a).[5] The entire content of consciousness makes up the content of the feeling component, and this secures the experiential aspect of emotions (Figure 1, Panel a). The father in our example may become conscious of the fact that something is wrong, that he has an urge to be aggressive, that his blood is boiling, that he is frowning, and that he has started to yell. Feelings can be labeled with an emotion word, but they can also remain unlabeled. Thus, the father may or may not label his feelings as anger. If he does, the label becomes part of his feelings and may color and shape them further.

It is worth noting that although the GDT can accommodate all the mental and bodily components typically considered to be part of emotional episodes, it also naturally accounts for episodes in which one or more of these components are absent. Take the absence of the feeling component. The GDT can easily explain cases in which a person displays "emotional behavior"—intense, disproportional shouting, for instance—while denying feeling angry. Here, all the steps in the goal-directed cycle are present, but few traces reach awareness, or the person has no desire to use any label.

Or take the absence of bodily components. At least five scenarios fit with the GDT. A first scenario is that the individual selects accommodation or immunization (i.e., two forms of covert or mental behaviors) rather than assimilation (i.e., overt behavior). A second scenario is that the individual gets stuck in the strategy/behavior selection stage because the decision is too difficult or because they do not see an appropriate strategy/behavior. A third scenario is that the individual actively selects a "passive behavior" option (i.e., do nothing) because they estimate this to have the highest expected utility. A fifth scenario is that a discrepancy gets resolved (with or without the person's own doing), so that no further action is needed. In the latter case, however, behavior may not be completely absent, for the individual may start exploring opportunities for the satisfaction of other goals (Moors, 2022, 2025). This may manifest in behavior with no other purpose than to increase the likelihood of detecting these opportunities, such as jumping up and down (see Fredrickson's, 2004, notion of broaden-and-build). The individual may also adopt the goal of communicating their success to others, which may manifest in attention-drawing or status-displaying behavior such as expanding the chest and raising the arms (Fridlund, 1994; Tracy & Matsumoto, 2008).

Valence

Valence can be regarded as a property of the feeling component. If we consider it as a dimension ranging from very negative to very positive, two aspects can be distinguished: a polarity or direction (whether it is positive or negative) and an extremity or intensity (more or less positive or negative; i.e., distance from the neutral mid-point). The GDT has made a first pass at formulating an account of the positive or negative valence of emotions (Moors, 2025; Moors et al., 2021), which remains open to empirical refinement and revision. In brief, the GDT states that the presence of an (actual or anticipated) stimulus-goal discrepancy (i.e., mismatch) gives rise to negative affect whereas the (actual or anticipated) absence of a stimulus-goal discrepancy (i.e., match) gives rise to positive affect.

Affect is not only generated in the discrepancy detection stage, but also in the behavior selection stage (Moors, 2022). After detecting a discrepancy, which generates negative affect at first, the individual may discover an effective behavior option to reduce this discrepancy, which generates positive affect in turn. It may be noted that the discovery of an effective behavior option to reduce a current discrepancy signals that the discrepancy will be effectively reduced in the future, which aligns with the case of positive affect generated by an anticipated discrepancy reduction. The upshot is that affect may dynamically change within and across goal-directed cycles.

Affect may not only stem from the *presence* or *absence* of a discrepancy but also from *changes* in discrepancies across two time points. Thus, positive affect not only stems from the (actual or anticipated) *absence* of a discrepancy but also from an (actual or anticipated) *resolution* or *reduction* of a discrepancy (Moors, 2023, 2025). Likewise, negative affect is not restricted to the *presence* of a discrepancy but also to an *increase* in discrepancy.[6] In other words, there may be positivity/negativity in getting closer to/farther away from a goal.

While the presence/absence or increase/decrease of a (current/anticipated) stimulus-goal discrepancy are factors that determine the negative/positive *polarity* of affect, several other factors are required to make proper sense of the *intensity* of affect. A first factor is the *magnitude* of the stimulus-goal discrepancy—given that such a discrepancy is present—which depends in turn on the content of the goal and the content of the stimulus representation. To illustrate, wanting to have 5 euros and receiving 1 euro will lead to more intense negative affect than wanting to have 5 euros and receiving 4 euros. Another factor is the *value* of the goal at stake. To illustrate, wanting to have 5 euros and receiving 1 euro will lead to more negative affect for an individual who cares more about money. In cases where a discrepancy is completely absent (i.e., a perfect match), the magnitude of the discrepancy is zero, but the value of the goal can still produce a positive affect.[7] In this way, the intensity of positive affect from closing the gap with the goal to have 5 euros will be proportional to the value attached to this goal.

Another factor, especially relevant for *anticipated* matches/mismatches, is the uncertainty or *probability* with which the matches/mismatches are expected. Thus,

anticipating the event of having 5 euros (i.e., anticipated match) may generate more positive affect when this event is more certain, and hence when the match is more certain. Once a match/mismatch with a goal is obtained, however, *unexpectedness* seems to boost the intensity of the resulting positive/negative affect (e.g., Szűcs et al., 2025, but see Marshall & Brown, 2006): un/expected matches with goals produce more negative/positive affect than expected ones. This observation raises the question whether (absolute/relative and current/anticipated) matches/mismatches with *goals* are sufficient to account for all the to-be-explained variety of affect or whether (absolute/relative and current/anticipated) matches/mismatches with *expectations* (i.e., expectedness/unexpectedness) must be added as a necessary determinant of affect (see discussion in Moors et al., 2021).

Because individuals pursue multiple goals, mis/matches with each of the activated goals constitute different sources of affect, generating different traces of affect that have to be combined in some way. This plausibly occurs in an additive fashion (i.e., affect traces accumulate), though the relationship between each source and affect may not be linear (i.e., not be strictly proportional), and different affect traces may carry different weights (i.e., some may have a greater effect on the total affect than others). The weight of an affect trace may depend on the salience of its source as well as on regulatory attempts, such that some traces become foregrounded while others recede into the background (Lambie & Marcel, 2002; Moors, 2022, p. 54). When this occurs in an alternating fashion, affect traces may not be fully blended but form heterogeneous mixtures.

Intensity, Urgency, and Automaticity

In the previous section, I presented valence as a dimension of the feeling component, itself characterized by polarity and intensity. The term intensity, however, has several other meanings. One meaning, already touched upon, is intensity understood as salience or the activation level of mental representations (e.g., goals, stimulus representations, and action tendencies). Another meaning—and the one under scrutiny in this section—is the intensity of behavior-related components: (a) the intensity of action tendencies, which aligns with the effort a person is willing to invest in a behavior, (b) the intensity of physiological responses, known as arousal, and (c) the intensity of the ensuing (subtle and coarse) behavior.

The GDT proposes that the high behavior-related intensity of episodes that people tend to call emotions has its origin in the high value of the first goal at stake in the goal-directed cycle. This high value also accounts for the urgency and automaticity often ascribed to emotions. Compare the situation of dropping a baby versus dropping an egg on the floor. The goal to keep the baby intact has a higher value than the goal to keep the egg intact. Because of this higher value, the action tendency to pick up the baby will be more urgent than to pick up the egg, which in turn, will lead to a more acute mobilization of the body and behavior, with a faster onset, a steeper rise in intensity, and a higher overall intensity (Verduyn et al., 2009).[8] In the less extreme cases of the father yelling at his child to restore his power versus to enforce obedience, the former goal may carry a higher value than the latter. This difference may account for the "emotional" tone of the former and the "instrumental" tone of the latter.

The value of the goal at stake is likely not the only factor that contributes to the intensity and/or urgency of the behavior-related components. In fact, all factors discussed as contributing to the intensity of valence may play a role here as well: the magnitude of the discrepancy (partly related to the value of the goal), the uncertainty of anticipated stimuli (in goal-directed cycles triggered by such stimuli), and the expectancy of the selected action option (in the action selection stage). The exact weighting and interaction of these factors are food for empirical research (ideally aided by computational modeling), though much can already be gleaned from research on effort investment (e.g., Neal et al., 2017). With enough imagination, the GDT can also make sense of cases in which the intensities of the three behavior-related components do not match, such as when motivational and somatic components are very intense while overt behavior is absent. If a person decides to suppress an action tendency that already generated bodily activity, this activity cannot be channeled into behavior, thereby causing a temporary overflow.

Irrationality

As discussed above, goal-directed processes are more likely to produce rational behavior than stimulus-driven ones because they can flexibly adapt to changes in the outcomes of behavior. This begs the question: If emotional behavior is indeed caused by a goal-directed process, as the GDT claims, how can the theory explain that this behavior often presents itself as irrational? The answer of the GDT is threefold: (1) there may be hidden goals; (2) there may be biases or mistakes; and (3) truth assent is not necessary (Moors, 2022, 2025). I elaborate on each in turn.

Hidden Goals. Some so-called emotional behaviors merely *seem* irrational, while they could in fact be rational. An agent has multiple goals, some of which may be more obvious and others more hidden, either from the perspective of the agent (i.e., more unconscious or less accessible) or from the perspective of external observers. A behavior option may have a high expectancy for a hidden goal but a low expectancy for an obvious goal. If the total expected utility of this behavior option is higher than that of other behavior options or of doing nothing, the behavior is still rational, while appearing irrational. Let us return to the father yelling at his child. This behavior may be costly in

the sense that it can jeopardize the relationship with the child, and therefore be judged irrational—and therefore emotional. On the other hand, the father may estimate that it is the best or even only behavior at his disposal for restoring his power, and this may have priority over the quality of the relationship, at least in the moment. The goal of exerting power may be somewhat hidden because it may not be the most virtuous goal to have. A more subtle scenario would be that the father is, in fact, aware of his goal to remain in charge, but not of the excessively high value he attaches to it.

Biases or Mistakes. The GDT can also account for emotional behavior that is genuinely irrational. It does so by invoking the possibility that agents hold biases or make mistakes in one or more steps of the goal-directed cycle. The representations in the goal-directed cycle are vulnerable to deviations from what is objective or normative.[9] Biases may have to do with the content of representations, such as deflated or inflated estimations of values of goals, likelihoods of anticipated stimuli, stimulus-goal discrepancies, and expectancies of behavior options. But biases may also relate to the activation level or salience of these representations (Köster et al., 2021). Some stimuli or goals may receive more attention than they deserve, while others get overlooked, and not all behavior options and their potential outcomes are always considered.

To illustrate biases in content or magnitude, the father may attach too much value to the goal to keep his power intact, or he may overestimate the discrepancy that the child's behavior constitutes with this goal. He may also overestimate the expectancy that yelling will help restore the perceived breach in power and/or underestimate the expectancy that yelling will create future discrepancies with other goals (e.g., the goal to maintain a good relationship with his child). In addition to overestimating the benefits and underestimating the costs of yelling, the father may also underestimate the benefits and overestimate the costs of alternative behaviors such as lecturing or calm disapproval.

To illustrate biases in activation level or salience, the alternative behavior of calm disapproval may simply not figure in the father's behavioral repertoire, or it may fail to reach the requisite activation level to be included in the equation. Likewise, some of the relevant benefits and costs of the considered behaviors may be unknown to the father, or they may fail to come to mind when they matter. Thus, the father may (temporarily) be blind to the harm that yelling creates for the relationship or to the benefits of remaining calm.

An important corollary assumption in the GDT is that biases and mistakes are more likely to occur when operating conditions are poor. An agent under time pressure or whose attentional resources are depleted, for instance, will not carefully weigh up all possible behavior options and outcomes and is more likely to select the behavior option from their repertoire that is most accessible. This will often be overtrained behavior that used to be effective or adaptive in the past but no longer is today (Buabang et al., 2023).[10] This corollary assumption helps understand why irrational behavior is observed more under poor operating conditions, such as a lack of time, attention, and motivation due to fatigue or stress (Edland & Svenson, 1993).

Such poor conditions may also be naturally characteristic of emotional episodes. The urgency in these episodes, stemming from highly valued goals and large discrepancies, may lead to self-imposed time pressure on the strategy/behavior selection process. The father's overvalued goal to stay in charge may lead him to perceive a small window for action, encouraging the hasty selection of his usual yelling spree and foregoing the consideration of other options. Thus, urgency is a double-edged sword. It not only intensifies the chosen behavior (as suggested in the section on intensity) but also narrows down the agent's action radius. Moreover, the high value of the goal in emotional episodes may also affect the salience of this goal relative to that of other information.[11] Thus, the father may be so consumed by attempting to reassert his dominance that he neglects the long-term costs of his behavior.

Truth Assent Not Necessary. Another way in which the GDT deals with cases of genuinely irrational behavior in emotional episodes is to argue that all representations in a goal-directed cycle, once encoded in memory, can be reactivated without an agent assenting to their truth. Indeed, stimuli, discrepancies, and expectations/expectancies may be entertained without believing that they are real or will become true. The mere activation of a representation may be sufficient to activate the next representations in the cycle, at least if this activation exceeds a certain threshold. Note that in addition to the reactivation of previously encoded representations, it is also possible that novel representations get formed and thereby activated on the spot without truth assent. Further note that the expression "without truth assent" should be qualified as "without *active* truth assent." This qualification is needed to make room for the possibility that *passive* truth acceptance is the default, and that it is the rejection of truth that requires a separate process of truth evaluation (Moors, 2022, p. 222).

Activation without active truth assent explains why emotions may be elicited by memories, imaginings, fiction, and virtual reality. For instance, remembering a social blunder may reactivate the discrepancy with the goal for social approval, in turn reactivating the goal to reduce this discrepancy by selecting the tendency to hide from sight. Likewise, imagining, or virtually experiencing (in film or virtual reality) falling from the stairs may activate the discrepancy with the goal for physical integrity, in turn activating the goal to reduce this discrepancy by selecting the tendency to protect one's vital body parts. The mere activation of an action tendency will translate into its implementation, supported by the

appropriate bodily activity, unless it is overridden by other action tendencies stemming from other cycles. The strength of the (re)activation of the action tendency should therefore be a critical factor in determining whether actual behavior does or does not occur. Moreover, the strength of the (re)activation of any component should be a critical factor in determining whether the content of the component surfaces in awareness and, therefore, colors the person's feelings. This helps understand why vivid perceptual inputs are often powerful in hijacking goal-directed cycles set up to deal with real discrepancies in daily life. Box 2 zooms in on how these three sources of apparent irrationality can make sense of two notoriously "though" cases of irrationality: recalcitrant emotions and emotion-driven arational behavior.[12]

Box 2. Recalcitrant emotions and emotion-driven arational behavior.

An emotion is recalcitrant when the stimulus evaluation presumably generating the emotion seems to conflict with the stimulus evaluation figuring in the agent's judgment. Examples are instances in which a person fears a stimulus while denying that it is dangerous, as in cases of fear of a house spider, fear of flying, and fear of heights in virtual reality. On first blush, cases of recalcitrant fear undermine the GDT because they would involve the co-existence of two conflicting evaluations in the discrepancy detection stage: the presence and absence of an anticipated discrepancy with the goal for safety. The approach taken by the GDT is to argue that the conflict is due to two different goal-directed cycles running in parallel, each at the service of a different goal and each involving a different action tendency or behavior. The cycle underlying "fear" involves an anticipated discrepancy with the goal for safety or some other goal, and may lead to the selection of fleeing, freezing, or any other behavior. The cycle underlying the (public or private) judgment that the situation is not dangerous is a cycle in which this judgment was generated as the best behavior option to reduce the discrepancy of a current state and the goal to produce an accurate or acceptable judgment. While the *fear cycle* may be at the service of a safety goal or any other goal, the *judgment cycle* may be at the service of epistemic or social goals. In some cases of recalcitrance, the goal at stake in the fear cycle is more hidden than that in the judgment cycle. In other cases of recalcitrance, the truth of the evaluations in the fear cycle is less endorsed than in the judgment cycle. But there may be more to it.

In the case of fear of a house-spider, the fear cycle may involve an anticipated discrepancy with the goal to keep unpredictably moving creatures away from one's body, whereas in the judgement cycle, the covert behavior of judging that the spider is not dangerous (i.e., that the spider will not produce a discrepancy with the goal for safety) is selected because it is most effective in satisfying an epistemic goal. Here, the crucial difference is that the goal involved in the fear cycle (i.e., to keep unpredictably moving creatures at bay) is not the same as the goal figuring in the judgment of the judgment cycle (i.e., to be safe or survive). Note, however, that if the goal underlying the fear cycle is hidden but the expected utility calculations are accurate, the cycle may appear irrational without actually being so.

In the case of fear of flying, the goal for safety in the fear cycle is the same as the one figuring in the judgment of the judgment cycle. In the fear cycle, the agent estimates the likelihood that the current state (being in a plane) will lead to a future state (being in a plane crash) that is discrepant with safety in the strongest sense. In the judgment cycle, the judgment about the danger of flying relates to safety in this strong sense. Here, the crucial difference is that the thresholds for the detection of an anticipated discrepancy with this goal in the fear cycle may be calibrated lower than the threshold for judgment about such a discrepancy in the judgment cycle. The fear cycle answers to the standard of survival, where the lower criterion for "detecting danger" stems from a liberal bias (i.e., a preference for avoiding misses rather than avoiding false alarms) in line with the principle "better safe than sorry." The judgment cycle, on the other hand, answers to the standard of accuracy, where the higher criterion for "detecting danger" stems from a conservative bias (i.e., a preference for avoiding false alarms over avoiding misses). Note that it may be questioned whether these "biases" are, in fact, biases. A liberal bias may be practically rational, whereas a conservative bias may be theoretically rational (see also Goffin, 2021).

What about the case of fear of heights in virtual reality, where individuals exhibit palpable bodily responses and the tendency to avoid jumping off a virtual cliff? In this case, there is zero chance that an anticipated discrepancy with the goal for safety will ever materialize. Thus, lowering the threshold for discrepancy detection (as in the fear of flying case) does not offer a solution here. I nevertheless think we can begin to understand this fear if we accept that vivid perceptual input can hijack goal-directed cycles that are in place to deal with heights in reality. In addition, the resistance to jump may stem from a real anticipated discrepancy with the goal to maintain a sense of control (see Moors, 2025).

Emotion-driven arational behavior has been invoked as a type of behavior for which the agent has no reason other than to satisfy the action tendency characteristic of an emotion (Scarantino & Nielsen, 2015). Examples are slamming the door out of anger and stabbing an opponent

who is already dead out of rage. Emotion-driven arational behavior is characterized as not goal-directed because it appears to serve no goal beyond enacting the behavior itself (here, the tendency to aggress). Again, the main approach taken by the GDT is to point at the existence of different (more or less hidden) goals, such as the goal to communicate or directly influence other people, even in cases where these people are only present in the person's head. Thus, an agent may slam the door out of anger to communicate that they are not to be taken for granted (see Moors, 2022, 2025).

Emotion Types

Goal-directed cycles present qualitative variety due to factors such as the presence(increase)/absence(decrease) of actual/anticipated discrepancies and the contents of the first and third goals at stake in the cycle. They also present quantitative variety due to factors such as the magnitude and salience of discrepancies, goal values, and response-outcome expectancies. All these factors may contribute to the emotion words that people use to label the episodes they find themselves in or observe in others. The rules for assigning emotion labels will largely be shaped by cultural norms and individual learning histories. To illustrate, the presence/absence of stimulus-goal discrepancies may lead to ascriptions of positive/negative emotions. Current discrepancies may be tied to anger or annoyance, whereas anticipated discrepancies may be tied to fear. Anger labels may be more likely when goals such as status and possession are involved, and fear labels when a safety goal is involved. Sadness may be used for discrepancies that cannot be reduced via assimilation (i.e., behavior) and that eventually (perhaps after immunization in the form of denial has failed) drive the agent towards accommodation, for instance, by replacing the goal to be physically close to an object with the goal to be close in spirit. Some people may rely more on the content of the third goal (i.e., action tendency) to guide the self-ascription of emotions: anger when they feel the urge to aggress, and no emotion when they find less costly ways to get what they want. Finally, the intensity of valence and of the behavior-related components may influence the use of emotion words denoting different intensities within the same "family" of emotion labels, as in annoyance, anger, and rage. Taken together, the GDT seems well-equipped to account for a wide range of vernacular emotion types.

Implications of the GDT

The GDT proposes that the same goal-directed mechanism underlies episodes that people call emotional and those they call non-emotional. Many apparent differences between the two types of episodes can be traced back to differences in the values of the goals at stake and/or in the magnitudes of the discrepancies, creating more urgency and intensity in valence and behavioral components. This is captured in the catch phrase "emotions as high-impact decisions." Intensity is probably not the whole story, if we consider that instrumental behavior can be very intense too. This suggests that some episodes may get labeled as emotional when discrepancy detection does not smoothly transition into a behavioral (or other) solution but creates some kind of breakdown in functioning (see Young, 1949; but see Leeper, 1948; Scherer, 2000). Setting the issue of labeling aside, the GDT insists that emotional and non-emotional episodes share the same underlying machinery. One key implication is that this robs us of a mechanistic basis for setting apart emotional episodes (or emotions) as a scientific set. The GDT can be regarded as eliminativist in this sense, even though it acknowledges that the various phenomena people call emotions may differ descriptively from those they are less inclined to call emotions.

A second implication of considering emotions to be high-impact decisions is that research into the influence of emotions on other things—such as perception, attention, memory, judgments, and decision making—could be reframed as research into the influence of high-impact decisions on these other things (Moors, 2022, pp. 236–237). For instance, research on the influence of (incidental) emotions on decision-making could be reframed as research into the influence of one decision-making process on another one. More specifically, this research would examine the influence of the proximal determinants of the former decision-making process (discrepancies, values, and expectancies) on the output (behavior) of the latter (see Moors & Fischer, 2019).

A third implication of recognizing the goal-directed nature of things called emotions is that it provides more handles for people suffering from maladaptive emotions. From the perspective of traditional, reactive theories, people are at the mercy of emotion regulation attempts when faced with the ruthless eruption of emotions from their reptilian brain. Such attempts are often doomed to fail, however, given the presumed automatic nature of the stimulus-driven processes generating emotions and the presumed non-automatic nature of the goal-directed process involved in emotion regulation. The GDT, on the other hand, does not focus on emotions as the target of regulation, but rather on representations of stimuli, goals, discrepancies, and expectancies, which are arguably more tangible and malleable (see e.g., Bogaert et al., 2026).

A final implication of the GDT is that it alters the debate on the moral responsibility of emotions. While reactive theories lean toward a passive view of the individual, naturally absolving them of responsibility for their initial (non-regulated) emotions, the GDT increases the space of responsibility in an undeniable way. The person may not choose whether to detect a discrepancy, but there is agency in the strategy/behavior option they select to reduce this

discrepancy. Moreover, keeping in mind the hierarchical architecture of the GDT, not only behavior options (i.e., third goals) but also subordinate goals (i.e., first goals) get selected at some point. These arguments notwithstanding, the individual's agency should not be overstated, not even in the GDT. First, people do not choose the superordinate, fundamental goals with which they are thrown into the world (e.g., survival, safety, and connectedness). Second, the subordinate goals they develop early in life to reach these superordinate goals depend on the reward structure of the external environment. Third, the selection of subordinate goals and strategies/behaviors may occur below the surface of awareness, thereby precluding thick agency. Insight into the workings of goal-directed processes might nevertheless offer a useful opening for increased agency and change (see also Solomon, 1973).

Acknowledgements

Preparation of this article was supported by Grant C14/23/062 of the Research Fund of KU Leuven and Grant G0AQO25N of the Fund for Scientific Research Flanders (FWO). I wish to thank Tamás Szűcs for useful discussions on valence.

Declaration of Conflicting Interests

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

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Fonds Wetenschappelijk Onderzoek, Onderzoeksraad, KU Leuven (grant numbers G0AQO25N and C14/23/062).

ORCID iD

Agnes Moors https://orcid.org/0000-0002-5137-557X

Notes

1. I capitalize Intentionality in the sense of "directedness" to separate it from intentionality in the common sense of "caused by an intention," following Searle (1983).
2. One peculiarity to be kept in mind is that this mechanism or information process is often considered to be part of the emotion or emotional episode itself, thereby blurring the distinction between constitutive and causal-mechanistic explanations (Moors, 2022, p. 48).
3. The list of basic emotions is not fixed, however, and contemporary evolutionary theories are progressively expanding it (see Keltner et al., 2019).
4. Panksepp (2012) does not rely on language to determine his set of basic emotions (i.e., fear, rage, panic/grief, lust, care, play, and seeking) but instead draws on existing evidence for neural circuits in animals. However, the question remains as to how the instincts to study were initially selected.
5. I proposed that the activation level of a representation (roughly aligning with the terms representational quality or salience) is the proximal determinant of the unconscious and conscious processing of this representation. This factor is itself fed by a number of remote factors such as time, intention, attention, recency, frequency, goal relevance, un/expectedness, and so on (Moors, 2016). Moreover, I proposed that the contributions of these factors to the total activation level combine in an additive way, which entails that the low quantity of one factor can be compensated for by a higher quantity of other factors. For instance, if time is limited, increased attention can still lift the activation level of a representation above a certain threshold.
6. An issue of debate is whether the mere absence/presence of a discrepancy is sufficient to generate affect or whether a change in discrepancy (reduction/increase) is necessary. Can people feel good or bad about what they have or lack or is affect only palpable when a change is detected or anticipated? The empirical study of this issue, however, is complicated by the fact that every *reduction* in the discrepancy with a static goal (e.g., the goal to be loved) can be considered as the *absence* of a discrepancy with the dynamic goal of obtaining this reduction (e.g., the goal of making progress towards the goal to be loved). Likewise, every *increase* in the discrepancy with one static goal can be considered as the *presence* of a discrepancy with the dynamic goal of being spared from such an increase.
7. For goals that are easily quantifiable, such as the goal to have a fixed amount of money, it makes sense to distinguish between the content and the value of the goal. However, such a distinction is less straightforward for goals that are harder to quantify. For instance, it may not be meaningful to distinguish between the amount of love one wants and the value one attaches to the goal of being loved. As a consequence, it may not be meaningful to distinguish between attaching a *high* value to being loved *a little* versus attaching a *low* value to being loved *a lot*. This is why I have neglected the distinction between the content and the value of goals in the past, sometimes stating that discrepancy is constituted by the value (instead of the content) of the goal and the stimulus.
8. This increased intensity may be warranted, since bigger problems often require bigger efforts to solve them, but it may also reflect a bias, known as overshooting. Research has shown that overshooting is more likely when the goal at stake has a higher value, as in the illustration of a football player who kicks too hard for the decisive penalty (Oudiette et al., 2019).
9. Note that only deviations from objective information, but not deviations from norms, can literally be called mistakes. In the case of genuine mistakes, theoretical irrationality is the origin of practical rationality.
10. We have argued that overtraining does not lead to the stamping in of a habit, with the [S:R-O^v] link transforming into an [S-R] link. Instead, overtraining should result in a very salient [S:R-O^v] link that defeats any newly established but correct [S:R-O^v] links, especially under poor operating conditions.
11. Although the content and salience of representations are conceptually distinct, crosstalk between them seems likely.
12. Arational behavior can be subsumed under the class of irrational behaviors because engaging in behavior for no reason still costs energy and hence goes against the goal to save energy (Moors, 2022, p. 43).

References

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Bogaert, L., Moors, A., & Raes, F. (2026). A goal-directed perspective on dampening of positive affect. *Psychological Review*.

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

Eder, A. B. (2023). A perceptual control theory of emotional action. *Cognition and Emotion*, *37*(7), 1167–1184. https://doi.org/10.1080/02699931.2023.2265234

Edland, A., & Svenson, O. (1993). Judgment and decision making under time pressure: Studies and findings. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgment and decision making* (pp. 27–40). Springer.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*, *3*(1), 34–38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

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

Fischer, M., Fini, C., Brass, M., & Moors, A. (2020a). Early approach and avoidance tendencies can be goal-directed: Support from a transcranial magnetic stimulation study. *Cognitive, Affective, and Behavioral Neuroscience*, *20*, 648–657. https://doi.org/10.3758/s13415-020-00793-6

Fischer, M., Kuppens, P., & Moors, A. (2020b). When socially excluded people prefer moralizing to anti- and prosocial behavior: Support for a goal-directed account. *Motivation and Emotion*, *44*, 508–524. https://doi.org/10.1007/s11031-020-09826-6

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359*(1449), 1367–1377. https://doi.org/10.1098/rstb.2004.1512

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

Goffin, K. (2021). Better scared than sorry: The pragmatic account of emotional representation. *Erkenntnis*, *88*(6), 2633–2650. https://doi.org/10.1007/s10670-021-00470-1

Heyes, C., & Dickinson, A. (1990). The intentionality of animal action. *Mind & Language*, *5*(1), 87–103. https://doi.org/10.1111/j.1468-0017.1990.tb00154.x

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

Köster, M., Moors, A., De Houwer, J., Ross-Hellauer, T., Van Nieuwerburgh, I., & Verbruggen, F. (2021). Behavioral reluctance in adopting open access publishing: Insights from a goal-directed perspective. *Frontiers in Psychology*, *12*, 1111. https://doi.org/10.3389/fpsyg.2021.649915

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

Lambie, J., & Marcel, A. (2002). Consciousness and emotion experience: A theoretical framework. *Psychological Review*, *109*(2), 219–259. https://doi.org/10.1037/0033-295x.109.2.219

Leeper, R. W. (1948). A motivational theory of emotion to replace "emotion as disorganized response". *Psychological Review*, *55*(1), 5–21. https://doi.org/10.1037/h0061922

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

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

Marshall, M., & Brown, J. (2006). Emotional reactions to achievement outcomes: Is it really best to expect the worst? *Cognition & Emotion*, *20*(1), 43–63. https://doi.org/10.1080/02699930500215116

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263–287. https://doi.org/10.1146/annurev-psych-122414-033550

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

Moors, A. (2023). Varieties of instrumental theories of emotional action: Commentary on "A perceptual control theory of emotional action". *Cognition and Emotion*, *37*, 1185–1192. https://doi.org/10.1080/02699931.2023.2266988

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. (2025). How the goal-directed theory of emotions can account for valence and irrationality: Reply to objections. *Acta Analytica*. https://doi.org/10.1007/s12136-025-00638-2

Moors, A., & Boddez, Y. (2017). Author reply: Emotional episodes are action episodes. *Emotion Review*, *9*(4), 353–354. https://doi.org/10.1177/1754073917692131

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

Moors, A., Fini, C., Everaert, T., Bardi, L., Bossuyt, E., Kuppens, P., & Brass, M. (2019). The role of stimulus-driven versus goal-directed processes in fight and flight tendencies measured with motor evoked potentials induced by transcranial magnetic stimulation. *PLoS One*, *14*, e0217266, 1–22. https://doi.org/10.1371/journal.pone.0217266

Moors, A., & Fischer, M. (2019). Demystifying the role of emotion in behaviour: Toward a goal-directed account. *Cognition and Emotion*, *33*(1), 94–100. https://doi.org/10.1080/02699931.2018.1510381

Moors, A., Van de Cruys, S., & Pourtois, G. (2021). Comparison of the determinants for positive and negative affect proposed by appraisal theories, goal-directed theories, and predictive processing theories. *Current Opinion in Behavioral Sciences*, *39*, 147–152. https://doi.org/10.1016/j.cobeha.2021.03.015

Neal, A., Ballard, T., & Vancouver, J. B. (2017). Dynamic self-regulation and multiple-goal pursuit. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(1), 401–423. https://doi.org/10.1146/annurev-orgpsych-032516-113156

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

Oudiette, D., Vinckier, F., Bioud, E., & Pessiglione, M. (2019). A Pavlovian account for paradoxical effects of motivation on controlling response vigour. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598–019-43936-7

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

Rolls, E. T. (2005). *Emotion explained*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570035.001.000

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., & Nielsen, M. (2015). Voodoo dolls and angry lions: How emotions explain arational actions. *Philosophical Studies*, *172*(11), 2975–2998. https://doi.org/10.1007/s11098–015-0452-y

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

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press. https://doi.org/10.1017/cbo9781139173452

Shiota, M. N. (2024). Basic and discrete emotion theories. In A. Scarantino (Ed.), *Emotion theory: The Routledge comprehensive guide* (pp. 310–330). Routledge.

Solomon, R. C. (1973). Emotions and choice. *The Review of Metaphysics*, *27*(1), 20–41. https://doi.org/10.1093/0195145496.003.0001

Szűcs, T., Wu, Y., Tuerlinckx, F., & Moors, A. (2025). Manuscript in preparation.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences*, *105*(33), 11655–11660. https://doi.org/10.1073/pnas.0802686105

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. https://doi.org/10.1177/1754073911410747

Verduyn, P., Van Mechelen, I., Tuerlinckx, F., Meers, K., & Van Coillie, H. (2009). Intensity profiles of emotional experience over time. *Cognition and Emotion*, *23*(7), 1427–1443. https://doi.org/10.1080/02699930902949031

Young, P. T. (1949). Emotion as disorganized response: A reply to professor Leeper. *Psychological Review*, *56*(4), 184–191. https://doi.org/10.1037/h0054176