Acta Analytica (2025) 41:187–201
https://doi.org/10.1007/s12136-025-00638-2

How the Goal-Directed Theory of Emotions Can Account for Valence and Irrationality: Reply to Objections

Agnes Moors[1]

Received: 14 January 2025 / Accepted: 13 March 2025 / Published online: 7 April 2025
© Springer Nature B.V. 2025

Abstract
In this reply paper, I address three objections that were raised against the goal-directed theory of emotions (Moors 2022). The first objection questions the appropriateness of aligning all emotions, including positive ones, with decision-making processes. The second objection challenges the theory's ability to account for the valence aspect of all positive and negative emotions, highlighting two sets of counterexamples. The third objection asserts that the theory fails to give a satisfactory account of the irrational flavor of certain emotions, in particular recalcitrant ones.

Keywords Emotion · Affect · Goal-directed · Irrationality · Recalcitrant

In this paper, I take the opportunity to reply to three objections that were raised by Pineda-Oliva (2025) against the goal-directed theory (GDT) of emotions (Moors, 2017a, 2017b, 2022, Chapter 7; Moors et al., 2017). The first objection questions how appropriate it is to align all emotions, including positive ones, with decision-making processes. The second objection is that the GDT cannot make sense of the valence aspect of certain emotions, in particular positive emotions with bodily responses such as pride and amusement and extreme positive and negative emotions such as euphoria and violent rage. The third objection is that the GDT does not give a satisfactory account of the irrationality of certain emotions, in particular recalcitrant ones.[1] Before discussing the three objections and my replies, it is worth presenting a brief sketch of the theory's basic elements in my own words.

[1] I organized the author's objections in a slightly different way to better clarify their connections.

✉ Agnes Moors
agnes.moors@kuleuven.be

[1] KU Leuven-University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium

1 Brief Sketch

The GDT proposes that the phenomena that people call emotions can be understood as goal-directed cycles of behavior causation or parts thereof. A cycle starts when a first goal (desired state; representation of a valued outcome: [O_1^v]) is compared with the representation of a current or anticipated state or stimulus ([S]). If the current/anticipated stimulus is discrepant, it activates a second goal ([O_2^v]), which is to reduce the current/anticipated discrepancy. This can be done by selecting one of three strategies: selecting and engaging in an action to bring the stimulus more in line with the goal (i.e., assimilation), replacing the first goal with another goal (i.e., accommodation), and changing the interpretation of the stimulus so that it appears less discrepant with the first goal (i.e., immunization). The selection of an action option in the case of assimilation is done by weighing up the expected utilities of the action options [R_1, R_2, …R_n] available in the agent's repertoire. The expected utility of one action option is the product of the value of the outcome (from the first goal) and the subjective probability or expectancy that the action will lead to this outcome given the stimulus ([S: R_1-O_1^v]). Once an action is selected, activation of the corresponding action tendency or intention, which is a third goal, takes place. This action tendency can be implemented in overt behavior. The actual outcome of this behavior is fed back to the cycle where it forms the input to a new run of the cycle. The cycle is repeated until the discrepancy is resolved or until another goal-directed cycle takes over. Note that the selection of the broad strategies of assimilation, accommodation, and immunization is also based on a weighing up of their expected utilities.

At least two forms of complexity must be added to this simple picture: People have multiple goals (i.e., horizontal complexity), which is why the expected utility of strategies/action options may be considered for several (active) goals. Differently put, strategies/action options may have several benefits and costs. In addition, goals are organized in a goal hierarchy (i.e., vertical complexity). This means that the first goal in a cycle may be a subordinate goal that was itself selected because it had the highest expectancy for reaching a superordinate goal. Likewise, the third goal in a cycle may give rise to the selection of ever more concrete subordinate goals downstream.

Aspects of all steps in the goal-directed cycle may enter into awareness, where they form the content of the agent's feelings or affect. These feelings can, but do not have to be labeled using emotion words (e.g., anger, fear, joy). An important feature of these feelings is valence, which I take to be a dimension with a polarity (positive or negative) and an extremity or intensity (distance from the neutral midpoint). The GDT proposes two sources of valence (Moors et al., 2021). The first source is situated in the discrepancy detection phase. Here, the presence of a (current or anticipated) discrepancy (i.e., mismatch) generates negative affect, with the magnitude of the discrepancy determining the intensity of this negative affect; the absence or reduction of the discrepancy (i.e., match) generates positive affect, with the value of the goal contributing to the intensity of this positive affect. The second source of valence is situated in the strategy/action selection

phase. Here, the presence of a strategy/action option with a high expectancy generates positive affect; the absence of such a strategy/action option yields negative affect. Note that the second source of affect could be reduced to the first source because when a strategy/action option with a high expectancy is present, then this amounts to the anticipation of the reduction or absence of a discrepancy.

The GDT holds that the episodes that people tend to call emotional vs. non-emotional share the same underlying mechanism (i.e., the goal-directed cycle described above) but differ in terms of the value of the first goal in the cycle. Higher values breed larger discrepancies and a stronger urgency to reduce these discrepancies. If assimilation is selected, this leads to more intense action tendencies, followed by more acute mobilization of bodily responses preparing for overt behavior. Given that the difference between emotional vs. non-emotional episodes is a matter of degree but not a matter of distinct underlying mechanisms, the theory suggests an eliminativist answer to the question of whether emotions qualify as a separate scientific set. With the key assumptions of the GDT established, I will now address the three objections pertaining to terminology, valence, and irrationality.

2 Terminology

In the GDT, the paradigmatic case of a positive emotion corresponds to a cycle in which the discrepancy detection phase yields the output that a stimulus-goal discrepancy is absent or resolved after which the goal-directed cycle is aborted. The author raises the question: If a goal-directed cycle is aborted and there is no strategy/action selection phase, should we call this a case of decision-making at all?

I think this issue can be put to bed as a terminological issue. I am not claiming that aborted goal-directed cycles count as full-blown decision processes, but rather that the GDT can make sense of the phenomena we call positive emotions, namely by presenting them as goal-directed cycles that only contain a discrepancy detection phase. Furthermore, I have employed the phrase "emotions as decision making" sparingly and rhetorically in writings, specifically in the context of negative emotions (e.g., Moors, 2022, p. 229), and as a provocative title for oral presentations. Thus, the author's shift from the thesis that emotions are decision-making processes (D-M) to the thesis that emotions are (complete or partial) goal-directed cycles (G-D-C) does not reflect any shift in my own thinking.

3 Valence

Terminological issues aside, the thesis that paradigmatic positive emotions involve a discrepancy detection phase without an action selection phase and subsequent action tendency appears incompatible with cases of positive emotions that do include an action tendency. As mentioned by the author, I have previously presented two solutions (Moors, 2022, p. 220). The first solution is that some positive emotions may involve the detection of an imminent or partial match that still requires some action from the agent in order to turn it into a complete match (e.g., the cake in front of you

must still be approached and consumed to produce a positive outcome). The second solution is that the complete match with one goal may open an opportunity to strive for the fulfilment of other goals, initially leading to a receptive attitude and exploratory behavior. The author thinks my solutions are not satisfactory, pointing at the counterexamples of (a) pride and amusement, presented as positive emotions that lack action tendencies while still involving bodily responses, and (b) extreme positive and negative emotions such as euphoria and violent rage.

3.1 Pride and Amusement

The author frames pride and amusement as emotions without action tendencies while still involving bodily responses. He does not specify what these bodily responses are, but it seems that the prototypical responses accompanying pride and amusement are in fact behavioral in nature, even if the behavior at stake may qualify as subtle (i.e., nonverbal "expressions") rather than coarse (i.e., full-body actions).[2] Pride has been associated with an expansive body posture (including an expanded chest and arms extended out from the body or raised; Tracy and Robbins, 2004) and amusement with laughter (Warren et al., 2021). Thus, pride and amusement may still involve the tendencies to engage in these behaviors and hence the need for bodily preparation. Of course, critics may now ask how the GDT can explain these action tendencies or behaviors.

In the case of pride, obtaining a match with one goal, such as winning a game, may create the opportunity for reaching another goal, namely signaling one's success to others, ultimately serving social approval or status. Adopting an expansive body posture may be selected as an effective way to signal one's success.

The case of amusement seems more complex. Laughter is a repetitive vocal pattern, portrayed as a semi-involuntary reflex elicited by sudden but pleasant shifts in psychological states (Morreall, 1982) or benign violations (Warren et al., 2021). The evolutionary story goes that laughter developed in early man to signal to others that a potential threat is not a concern after all (Warren et al., 2021). The GDT does not deny the existence of reflexes or instincts (i.e., innate stimulus-driven mechanisms), but proposes that once the behavior dictated by the instinct has been executed, organisms will start encoding the outcomes of this behavior. From then on, the behavior will also be governed by these outcomes and hence be goal-directed. Once laughter is reliably followed by positive social outcomes in certain contexts, people learn to select this behavior in these contexts to regain these social outcomes.

The upshot is that there may be cases of positive emotions, such as pride and amusement, where the action tendencies generating bodily responses are selected to reduce discrepancies with communicative or other social goals. To counter the potential objection that people also experience pride and amusement when they are alone, I will argue that they can simulate actions to communicate with the (perhaps

[2] Note that the terms subtle and coarse do not refer to the intensity but rather to the potential impact that the behavior has on the physical world. Although subtle behavior has a lower material impact, it should not be underestimated in terms of its social impact (Moors, 2022, p. 53).

ever-present) audience in their heads (Fridlund, 1994; Fridlund et al., 1990; Parkinson, 1995; see Moors, 2022, p. 286).

3.2 Euphoria and Violent Rage

Another set of counterexamples that allegedly defy the valence account of the GDT are extremely positive and negative emotions such as euphoria and violent rage. The author argues that the extremely positive emotion of euphoria cannot exist in the GDT because when one discrepancy is completely solved, new discrepancies should open up, creating negative affect that should tone down the initial positive affect. The author also argues that the GDT cannot account for the extremely negative emotion of violent rage, in which aggressive behavior is selected to reduce the discrepancy. Here, the initial negative affect from the discrepancy should be cancelled out by the positive affect stemming from the high expected utility of the aggressive behavior.

I will argue that the GDT is fully equipped to account for euphoria and violent rage, even if I have not explicitly applied the theory to these phenomena before. To do this in a convincing manner, I will start by systematically discussing eight assumptions of the theory that were either (partially) misrepresented or overlooked. After that, I will try to demonstrate how these assumptions can be exploited to help us understand euphoria and violent rage.

(1) The author is correct that the presence/absence of a stimulus-goal discrepancy (mismatch/match) determines the polarity of affect (positive/negative) and that the magnitude of a discrepancy or mismatch determines the intensity of negative affect. However, the statement that the magnitude of a match determines the intensity of positive affect is less precise (Pineda-Oliva, 2025). I have aligned the intensity of affect with appraisal theory's notion of goal relevance (Moors et al., 2021), and I have argued that goal relevance depends on the value of the goal at stake in a cycle (e.g., Moors, 2022, p. 226). In the case of a mismatch, the value of the goal and the stimulus co-determine the magnitude of the discrepancy and hence the intensity of negative affect. In the case of a (perfect) match, on the other hand, the magnitude of the match is always the same (it cannot vary) so that only the value of the goal (which can vary) is left to determine the intensity of positive affect. The upshot is that the value of the goal at stake should not be overlooked as an important determinant of the intensity of the valence of an emotion. Because the value of a goal allows for wide variety (from zero to extremely high) and absolute intensities of affect traces stemming from different sources are a priori unknown, it is hard to predict what the net resulting affect will be when traces get combined.

(2) The author is correct in stating that (a) affect may vary across the emotional episode and (b) affect stemming from different sources may be combined. However, he also seems to go beyond these two assumptions, making corollary assumptions that are not part of the GDT and that are unduly constraining. In particular, the author writes: "I will also be assuming that its valence [the valence of the

emotional episode] at a particular moment will be the net result of the amount of positive and negative valence contributed by different elements of emotion generating valence up to this point" (Pineda-Oliva, 2025). For one thing, it is not entirely clear whether the author assumes that the affect traces from different sources are combined in a simple sum or not, but the GDT does not make any claims about this. All traces of affect that are present at a particular time should be integrated in some way. This may plausibly be done in an additive way, but different sources of affect could receive different weights (see below, Assumptions 5 to 8) and they could even be combined in a nonlinear way. For another thing, the author seems to assume in this citation that the affect at any moment is the sum of *all* the previously encountered affect traces. Yet it is unknown how long a single affect trace lingers on, and therefore how long it can exert a diminishing or enhancing effect on the net affective state of an agent.

(3) Although the author acknowledges the assumption of the GDT that valence may not only stem from *current* matches and mismatches but also from *anticipated* matches and mismatches, he does not seem to fully embrace the implications of this assumption. For instance, the number of goals for which an agent may imagine future fulfillment is limitless, even if unrealistic.

(4) Some formulations of the GDT leave open the possibility that positive affect not only stems from an actual or anticipated *absence* of discrepancy but also from an actual or anticipated *resolution* or *reduction* of discrepancy (e.g., Moors, 2023, p. 1188; see also Bogaert et al., 2025). The latter involves a change across two time points going from a mismatch to a match (resolution) or from a larger to a smaller mismatch (reduction).[3] Thus, there may be positivity in getting closer to a goal.

(5) Several assumptions concern factors that may contribute to the weight of an (objective) affect source. In the case of anticipated matches or mismatches, uncertainty may be such a factor (as suggested by Moors et al., 2021, Footnote 6). This entails that the positive affect from an action option with an expectancy above zero but below one may be weighted by the remaining uncertainty.

(6) Further weights pertain to biases, which may arise in all steps of the goal-directed cycle. Examples are the over- or underestimation of the magnitudes of discrepancies, values, and expectancies (see Köster et al., 2021). Another bias worth contemplating is framing (Tversky & Kahneman, 1981). By definition, any partial match is a partial mismatch. Thus, the same distance between a stimulus and a goal could be framed either as a partial match or a partial mismatch, and this may influence the polarity of affect. To understand how this could happen, let us consider the next assumption.

[3] A change across time points involves a comparison, where the first time point serves as the reference for the second time point. Note that the GDT can be considered one type of reference value theory of affect, where the reference point is the (first) goal at stake in the cycle. The jury is not yet out on the issue whether a second reference point is required in the form of "discrepancy on a previous time point." This is because the *reduction* of a discrepancy with one primary goal (e.g., the goal to be close to someone) can also be conceptualized as the *absence* of a discrepancy with the meta-goal to reduce the discrepancy with the primary goal (e.g., the goal to make progress in getting close to someone).

(7) Biases may not only pertain to the content of representations but also to their quality, known as "representational quality" (Moors, 2016, 2022, p. 23; see also Köster et al., 2021). This term largely refers to the level of activation or salience of a representation. A representation's quality must exceed a first threshold to be processed unconsciously and be causally efficacious (i.e., to pass its activation on to subsequent representations in the cycle) and a second threshold to be processed consciously (i.e., to have its content enter into awareness). Consistent with the idea that rationality is bounded rather than Olympian, not all goals are constantly in an active state, which entails that the number of goal-directed cycles running simultaneously will be limited. Further, within a single goal-directed cycle, the action options considered will be limited as will be the number of potential outcomes of these action options. Affect, which I take to be conscious, preys on these representations. Consequently, not all objectively potential sources of affect will actually generate affect, and when they do, the representational quality of these sources may be an additional factor contributing to the weight of the affect. I proposed that representational quality is fed by a number of factors such as time, intention, attention, recency, frequency, and so on (Moors, 2016). While representational quality is a proximal factor contributing to the weight of affect traces, the feeding factors can be seen as remote factors contributing to this weight. But even after an affect trace was generated, the possibility remains that one trace gets foregrounded in attention at the expense of others (Lambie & Marcel, 2002; Moors, 2022, p. 54). This opens up the possibility that affect traces are not always perfectly blended (i.e., homogenous mixtures) but could remain discernible to some extent (i.e., heterogenous mixtures).

(8) A final element that may modify the contribution of certain sources of affect is affect regulation. Indeed, an agent may decide to upregulate certain traces of affect (e.g., by the strategies of savoring or reappraisal) while downregulating others (e.g., by the strategies of suppression or distraction[4]).[5]

Taken together, the eight assumptions spelled out above should already take us a long way in accounting for cases of extreme positive and negative emotions. Let us return to the extreme positive emotion of euphoria and consider its unfolding over time. When an agent detects a complete match with a very highly valued goal, the resulting positive affect should be very high to begin with. At this point, the agent may employ the affect regulation strategy to savor the affect by attentionally foregrounding it and blocking out anything else. If this is not tenable for long, the first discrepancy that pops up may pertain to the meta-goal of fulfilling a host of other, primary goals that remain unspecified at this stage. The discrepancy with

[4] Note that in the affect regulation strategy known as distraction, the agent has the goal to direct attention away from a source of affect. If one accepts that not all attention may be driven by goals (Moors, 2016), this affect regulation strategy does not overlap entirely with the role of attention discussed in the previous paragraph.

[5] Affect regulation counts as assimilation in the face of a discrepancy with a goal to experience positive (or negative) affect (Moors, 2022, p. 242). Another way to reduce this discrepancy is to turn to immunization, that is, downplaying one's current/anticipated affect (Bogaert et al., 2025).

the meta-goal may produce a slight trace of negative affect along with exploratory behavior and a receptive attitude. As long as the specific primary goals are not on the agent's radar, then neither may be the discrepancies with these goals. Once these discrepancies do become salient, they may indeed throw some additional negative affect into the mix. However, this negative affect may be quickly overshadowed or erased if the agent revels in the prospect that all these goals will be easy to fulfill, even if this is an overestimation. The complete match from before may even fuel this overestimation (e.g., through generalization). The resulting positive affect may again be boosted or prolonged through savoring.

As this scenario demonstrates, the GDT can explain the occurrence of euphoria. But the theory also explains why euphoria is rare and fleeting in most people's lives. People find it hard to enjoy success precisely because other worries tend to remain prominent or quickly resurface.

Let us now return to the extreme negative emotion of violent rage, where the negative affect stemming from the discrepancy allegedly would be cancelled out by the positive affect stemming from the high expected utility of aggressive behavior. My reply is three-fold. First, it may be questioned whether the expected utility of aggressive behavior is really that high. Aggressive behavior may have the highest expected utility among the action options available from an agent's repertoire at some point. However, this does not mean the expected utility is high in absolute terms, even from the agent's perspective. The agent may realize that lashing out in anger not only has benefits but also costs, but they may select it for lack of a better action option and because passivity seems worse. My second reply questions the extreme negative affect experienced in episodes of violent rage. Some research suggests that anger—and by implication, rage—is not experienced as extremely negative and sometimes even as positive (see reports of satisfaction, relief, and sadistic pleasure; Averill, 1982; Baumeister & Bushman, 2007; Bushman et al., 2001). My third reply capitalizes on the idea that affect traces may not combine into a homogeneous blend but may take the form of heterogeneous mixtures in which the edge of each trace is not softened by the other traces.

To connect the dots, I will now consider the unfolding of violent rage over time. When confronted with a huge discrepancy with a very highly valued goal, the agent may experience extreme negative affect. When selecting and acting out the aggressive behavior, the agent may actually experience some positive affect. If the expected utility of aggressive behavior is not excessive in absolute terms, however, this positive affect will hardly be sufficient to *cancel out* the negative affect (as the author claims).

Rather than combining into a seamless blend, however, the two traces of affect may also combine into a heterogeneous mixture, where the negative affect from the insult remains sharply discernible from the gratifying aspects of the aggression, resulting in reports of "mixed feelings." Perhaps the sustained energizing of the aggressive behavior not only requires the renewed selection of this behavior across cycles but also the continuous reactivation of the discrepancy, thereby inevitably also renewing the negative affect.

Another possibility is that the negative trace gets foregrounded at the expense of the positive trace (perhaps because foregrounding positive aspects of

aggression is itself discrepant with moral norms), resulting in reports of extreme negativity. If, after the peak of the aggression is over, the costs of the aggressive behavior start becoming more prominent, the negative affect may rise further.

Taking stock, cases that seem puzzling at first blush force us to systematically map out all the factors in the GDT that may contribute to the net total affective experience of an agent. To assess whether these factors successfully account for the variety of affective phenomena that people may experience in daily life, intuitive analyses alone may not be sufficient. Empirical research and simulations assisted by mathematical models will ultimately be required.

4 Irrationality

As indicated by the author, the GDT proposes three ways to explain the practically irrational flavor of the phenomena called emotions, the idea that under the influence of an emotion, people tend to act in ways that hampers goal satisfaction or well-being (Moors, 2022, p. 228). The first explanation is that some instances of emotional behavior *appear* irrational but are in fact rational. This is because an agent may have goals that are more or less hidden (unconscious or less accessible). While a behavior may hamper the agent's explicit goals (thereby appearing irrational), it may still further hidden goals with a higher value in that moment (thereby being rational).

The second explanation acknowledges that some emotional behaviors are genuinely irrational and attributes them to biases or mistakes in one or more steps of the goal-directed cycle. Biases may have to do with content, such as the over- or underestimation of the magnitudes of discrepancies, values, and expectancies. Biases may also relate to representational quality, such as when one goal is more salient than another one (independent of its value) and when certain action options or certain outcomes of these options are not considered (Köster et al., 2021). 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, for instance, will not carefully weigh up all possible action options and outcomes and is more likely to repeat overtrained behaviors that used to be effective in the past but no longer are today (Buabang et al., 2023).

The third explanation capitalizes on the assumption that the representations in a previously stored goal-directed cycle may be activated without an agent assenting to their truth (e.g., discrepancies and expectancies may be entertained without believing they are real) and this may be sufficient to activate the action tendency in that cycle.

The author spells out objections against these three explanations using instances of recalcitrant emotions as counterexamples. These include pathological cases such as spider phobia and the more mundane cases of fear of flying and stage fright. I will first discuss the objections and counterexamples relevant to the first two explanations before doing the same for the third explanation.

4.1 First Two Explanations: Hidden Goals and Biases or Mistakes

The author discusses spider phobics in treatment who have the tendency to withdraw from a spider in exposure therapy while acknowledging that spiders are not dangerous. Taking self-preservation to be a central, hidden goal in spider phobia, the author states that the first two explanations amount to arguing that spider phobics hold the mistaken unconscious belief that spiders are dangerous—i.e., discrepant with the goal for self-preservation—while simultaneously holding the conscious belief that they are not dangerous, or denying that they are dangerous to save face (an action at the service of an impression management goal). The author then dismisses these explanations on the grounds that it would be preposterous to claim that spider phobics still believe spiders to be dangerous, given that they judge their own fear as recalcitrant and seek therapeutic help. If we accept this conclusion, the author continues, then there is also no reason to assume that the spider phobic's withdrawal tendency is based on an overestimation of the expected utility of withdrawal behavior. Likewise, for fear of flying and stage fright, the author deems it unlikely that there is a level at which the agent believes they are in danger and/or that withdrawal or cautious behavior is the best choice.

Although I am not convinced that the spider phobic's own acknowledgment of recalcitrance and search for therapeutic help undermine the role of unconscious or less accessible beliefs that spiders are dangerous, my reply here will focus on the author's insistence that the central goal in spider phobia is self-preservation. More plausible goals may be the goal for bodily integrity, in the sense of keeping spiders away from bodily orifices (given their unpredictable movements; Lindner et al., 2019; Moors, 2022, p. 229) or in the sense of avoiding contamination (Matchett & Davey, 1991). The risks of spiders violating these goals may more plausibly be overestimated by spider phobics.

In fear of flying, the agent may arguably envisage a potential discrepancy with the goal of self-preservation. The agent's acknowledgment that plane crashes are rare, in the sense that they have a low probability across instances, does not protect them from getting unlucky in a single instance. And if they do get unlucky, the consequences are fatal.

Likewise, in stage fright, the musician may realize there is no guarantee that the performance will be smooth, and if they do mess up, their reputation will be ruined. To reduce this anticipated discrepancy, the musician may select the ineffective (covert) action option to monitor their performance at the expense of the more effective (covert) action option to focus on the music. This may result in the performance actually deteriorating, further encouraging monitoring, thereby entering into a vicious cycle (Moors & Fischer, 2019; Vancouver & Tischner, 2004). In all three recalcitrant fear cases (spider phobia, fear of flying, and stage fright), the goals and beliefs driving the agent's behavior may be less accessible than the declarative belief reported by the agent.

Let us now turn to the corollary assumption that mistakes are more likely when operating conditions are poor such as a lack of time. The author accepts this scenario as an explanation for non-emotional action slips, such as typing in an old, no longer valid password (see Buabang et al., 2023), but not as an explanation of irrational

emotional behavior. This is because the author believes that "one crucial element […] is missing, as the operating conditions [in emotional episodes] are not in general poor." (Pineda-Oliva, 2025). I beg to disagree. I have argued that compared to non-emotional episodes, the goal at stake in emotional episodes has a higher value, leading to a larger discrepancy, and more urgency to reduce this discrepancy (Moors, 2022, p. 230). This urgency may lead the agent to self-impose time pressure on the action selection phase, which may result in the selection of accessible but suboptimal action options. A case in point is when people dash out of a burning building instead of exiting calmly, thereby increasing the risk of dire consequences.

4.2 Third Explanation: Activation Without Truth Evaluation

The author also criticizes my third explanation (Moors, 2022, p. 229), which suggests that the presence of the spider may activate a goal-directed cycle in which the spider is dangerous, without assenting to the truth of the activated information. This can be framed as a contrast between believing and imagining, which the author argues is devoid of conflict, as in the case where one imagines being rich while believing that one is not.

While the third explanation may contribute to spider phobia, it is best illustrated when invoked to explain fear of heights when an agent is asked to jump off a cliff in a virtual reality room. Although the chance of actually falling from a height is zero, the chance of falling in the virtual world is 100 percent and this may hold the agent back from jumping. Note, moreover, that although the expected outcome of hitting the ground is fictional, the expected outcome of experiencing a loss of control while falling may be real.[6] The upshot is that the agent does imagine the (fictional and real) outcomes of their actions and that a goal-directed process is therefore at play. The vividness and value of the outcomes may be what distinguish this case from the case in which a person imagines being rich while believing that they are not: The goal to be rich in the money case may have a lower value than the goal for self-preservation in the cliff case, and the vividness of being rich in the money case (i.e., an imagined actual match) may be lower than the vividness of approaching or hitting the ground in the cliff case (i.e., an imagined anticipated mismatch). If the goal to be rich would be extremely highly valued and vivid, an agent might also get ecstatic when imagining being rich.

The author further maintains that the GDT is unable to explain the practical irrationality of emotions without taking recourse to a duality in mechanisms. In particular, he suggests that in order to account for the differences between emotional and non-emotional behavior, I have traded the duality between stimulus-driven and goal-directed mechanisms by the duality between associative and rule-based operations. Let me first reiterate that the GDT is in fact still a dual-process theory in the original sense in that it does not deny the existence of stimulus-driven processes (whether

[6] I am drawing from a personal experience, in which I was unable to step off the cliff until I was told that the descent would be very slow, ensuring that the visual input during the descent would not induce the sense of losing control.

innate or learned), even if it minimizes their explanatory territory. Crucially, the theory does not exploit the duality between stimulus-driven and goal-directed mechanisms to account for the duality between irrational emotional behaviors and rational non-emotional behaviors. Instead, it argues that most cases of irrational emotional behavior can be explained through goal-directed processes with hidden goals (first explanation), biases or mistakes (second explanation), or a lack of truth evaluation (third explanation).

Of course, if dissociations occur between what an agent thinks or feels, as in recalcitrant emotions, the GDT will have to invoke some kind of contrast or duality, but the duality need not be in terms of mechanisms or even operations. In the first explanation discussed above, the duality is a matter of different cycles at the service of different goals, of which one is more accessible than the other. The second explanation adds to this by assuming that the less accessible cycle may also suffer more from biases. But what about the third explanation? This explanation relies on the duality between the absence vs. presence of truth evaluation, and this may indeed be considered a duality in terms of mechanisms. Crucially, however, this duality does not map onto the dualities between stimulus-driven vs. goal-directed, associative vs. rule-based, or emotional vs. non-emotional. Let me clarify the relations between these four dualities.

Stimulus-driven processes, as commonly understood, consist in the activation (i.e., associative operation) of associations. Associations are understood as representations of unqualified relations between elements (S-R) that, by definition, cannot be truth-evaluated. Goal-directed processes consist in the computation (i.e., a complex operation, which need not be rule-based[7]) or activation (i.e., an associative operation) of representations. Several of these representations contain propositions, understood as qualified relations (i.e., R *leads to* O). Although qualified relations can be evaluated for truth, they need not be.

Evaluating the truth of propositions in goal-directed processes may be considered as an additional step, thereby requiring at least one extra mechanism. An alternative option is that the propositions are treated as true by default unless there is evidence to the contrary (Moors, 2022, p. 222). Here, only the evaluation of something as "not true" requires an additional mechanism. No matter which mechanistic story turns out to be correct, what matters is that the dichotomy between the presence vs. absence of truth evaluation does not align with the dichotomy between emotional vs. non-emotional. Everyday instrumental behavior may equally well be undergirded by goal-directed processes relying on associative operations for which no separate step of truth evaluation is required. So even if certain forms of recalcitrance do rely on a duality in mechanism, this duality does not ground the distinction between emotional vs. non-emotional phenomena.

[7] As I have argued elsewhere, the dichotomy between associative and rule-based is elusive and may more fruitfully be understood as a gradual distinction, in terms of degree of abstraction (Moors, 2014, 2022, p. 21).

5 Conclusion

In this reply paper, I responded to three objections raised against the GDT. After dealing with an initial objection about the proper use of the term decision making, I treated two substantial objections about the scope of the GDT. One contends that the GDT fails to adequately account for the valence aspect of emotions, while the other asserts that the theory struggles to explain the irrational flavor of emotions. Replies to the latter two objections rested on the strategies to reanalyze presumed counterexamples by (a) questioning the way in which some of these counterexamples were portrayed and (b) clarifying how the GDT can still make sense of them by correcting certain misinterpretations and highlighting implications that remained underexploited so far.

To address the valence objection, I reanalyzed the alleged counterexamples of pride, amusement, euphoria, and violent anger. I reframed pride and amusement as involving subtle behavior at the service of communicative or social goals. I explained euphoria and violent anger through the interplay of a collection of factors that had been overlooked in the target paper. It must be kept in mind, however, that several factors are unknown at this stage and await empirical determination (e.g., the absolute intensity of valence stemming from all mentioned sources; how long affect lingers on after it has been generated; how the affect traces from different sources combine; the sources' representational quality; which affect regulation strategies are operative and effective).

The large list of unknowns is compatible with the notion that the GDT, in its current form, should be regarded as a mechanistic sketch (Moors, 2022, p. 222). Further theoretical and empirical work is needed to determine these unknowns but also to examine whether the GDT must be extended by additional sources of affect (even if this would come at a cost of parsimony). This can be done with the help of simulations using mathematical models, and by testing these against human affect reports in empirical research. My colleagues and I have recently embarked on this challenge in our lab.

To address the irrationality objection, I reanalyzed tough cases of recalcitrant fear such as spider phobia, fear of flying, stage fright, and fear of heights in virtual reality, all characterized by a dissociation between the tendency to withdraw (physically or mentally) and the belief that the situation is not dangerous. To counter the objection that there is no level at which agents still believe they are in danger, I have proposed finetuning the precise content of the goals and the nature of the biases that might plausibly be at stake. I have further underlined that the high value of the goal at stake in emotional episodes breeds urgency, leading to self-imposed time pressure, thereby increasing the likelihood for mistakes. To counter the objection that the mere activation of goal-directed cycles is not a good model for recalcitrance because imagining one thing and believing another does not create a conflict, I have argued that imagining does have the potential to activate action tendencies if the outcomes have a high value and are vivid.

Let me close by briefly commenting on the author's own suggestion to return to innate stimulus-driven processes (affect programs) as the fallback solution

whenever the GDT faces a challenge. While it may seem convenient to invoke innate stimulus-driven processes (known as instincts or affect programs) to explain the irrationality of some emotions, such a strategy faces a few challenges of its own: First, convincing empirical support for emotion-specific stimulus-driven processes is lacking. Second, theories capitalizing on stimulus-driven processes do not give us principled criteria to distinguish between emotional and non-emotional stimulus-driven processes. Third, stimulus-driven processes fail as a model for emotional phenomena that matter most to us (e.g., in social interactions). Specifically, how stimulus-driven processes are supposed to account for emotions such as pride, amusement, euphoria, and violent rage is left to our imagination.

Funding The 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).

Declarations

Ethical Approval Not applicable.

Competing Interests The author declares no competing interests.

References

Averill, J. R. (1982). Anger and aggression: An essay on emotion. *Springer-Verlag*. https://doi.org/10.1007/978-1-4612-5743-1

Baumeister, R. F., & Bushman, B. J. (2007). Angry emotions and aggressive behaviors. In G. Steffgen & M. Gollwitzer (Eds.), *Emotions and aggressive behavior* (pp. 61–75). Hogrefe & Huber Publishers.

Bogaert, L., Moors, A., & Raes, F. (2025). A goal-directed perspective on dampening of positive affect. Manuscript submitted for publication.

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*, 496–508. https://doi.org/10.1037/xge0001280

Bushman, B. J., Baumeister, R. F., & Phillips, C. M. (2001). Do people aggress to improve their mood? Catharsis beliefs, affect regulation opportunity, and aggressive responding. *Journal of Personality and Social Psychology, 81*(1), 17–32. https://doi.org/10.1037/0022-3514.81.1.17

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

Fridlund, A. J., Sabini, J. P., Hedlund, L. E., Schaut, J. A., Shenker, J. I., & Knauer, M. J. (1990). Audience effects on solitary faces during imagery: Displaying to the people in your head. *Journal of Nonverbal Behavior, 14*(2), 113–137. https://doi.org/10.1007/bf01670438

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

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

Lindner, P., Miloff, A., Reuterskiöld, L., Andersson, G., & Carlbring, P. (2019). What is so frightening about spiders? Self-rated and self-disclosed impact of different characteristics and associations with phobia symptoms. *Scandinavian Journal of Psychology, 60*(1), 1–6. https://doi.org/10.1111/sjop.12508

Matchett, G., & Davey, G. C. L. (1991). A test of a disease-avoidance model of animal phobias. *Behaviour Research and Therapy, 29*(1), 91–94. https://doi.org/10.1016/S0005-7967(09)80011-9

Moors, A. (2014). Examining the mapping problem in dual process models. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), Dual process theories of the social mind (pp. 20–34). Guilford Press.

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*, 1–19. 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*, 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., & Fischer, M. (2019). Demystifying the role of emotion in behaviour: Toward a goal-directed account. *Cognition & Emotion, 33*, 94–100. https://doi.org/10.1080/02699931.2018.1510381

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/10.1177/1754073916666695

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

Morreall, J. (1982). A new theory of laughter. *Philosophical Studies*, *42*(2), 243–254. https://www.jstor.org/stable/4319555

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

Pineda-Oliva, D. (2025). Some reflections on the goal-directed theory of emotion. *Acta Analytica*.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15*(3), 194–197. https://doi.org/10.1111/j.0956-7976.2004.01503008.x

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Warren, C., Barsky, A., & McGraw, A. P. (2021). What makes things funny? An integrative review of the antecedents of laughter and amusement. *Personality and Social Psychology Review, 25*(1), 41–65. https://doi.org/10.1177/1088868320961909

Vancouver, J. B., & Tischner, E. C. (2004). The effect of feedback sign on task performance depends on self-concept discrepancies. *Journal of Applied Psychology, 89*(6), 1092–1098. https://doi.org/10.1037/0021-9010.89.6.1092

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.