SPECIAL ISSUE: MOOR SECTION

emotion review

Emotion Review
Vol. 0, No. 0 (February 2026) 1–5

ISSN 1754-0739
DOI: 10.1177/17540739261426001
https://journals.sagepub.com/home/emr

Moors' Eliminativism Cross-Examined: Why Emotions Are Not Goal-Directed Cycles

Andrea Scarantino
Department of Philosophy, Georgia State University, Atlanta, GA, USA

Abstract

Moors' eliminativist theory of emotions aims to show that understanding emotional behavior as goal-directed behavior when the goals are high value explains all that is worth explaining about behavior without invoking the concept of emotion. I argue that eliminating emotions in favor of goal-directed cycles has major explanatory costs, because emotional behavior differs in important ways from behavior governed by cost-benefit analysis. I compare and contrast Moors' theory with my own Motivational Theory of Emotions (MTE) with respect to two explanatory challenges in particular—emotion-induced decisional paralysis and recalcitrance. I conclude that we cannot make sense of these affective phenomena in purely goal-directed terms, and that a stimulus-driven process of behavior causation such as the one posited by MTE is required for explanatory purposes.

Keywords
emotion, Moors, recalcitrance, decision-making, Motivational Theory of Emotions, stimulus-driven processes, goal-directed processes

Moors' (2022) goal-directed theory of emotions is the most sophisticated eliminativist theory currently available. It aims to show that understanding emotional behavior as goal-directed behavior explains all that is worth explaining about behavior without invoking the concept of emotion. As Moors (2022, p. 225) puts it, "emotions are goal-directed cycles in which a discrepancy is [detected between a stimulus and a high value goal], and in which the person selects a physical or mental action based on a cost-benefit analysis, or no action if none can be found." Once the scientific concept of goal-directed cycle is on board, emotion concepts can be purged from affective science with no explanatory loss.

I find Moors' (2022) eliminativist project in equal parts daring, brilliant, and misguided, because emotions are not goal-directed cycles, although they crucially involve goal-directed decision-making. Nevertheless, there is much to learn from Moors' (2022) unapologetic attempt to eliminate emotions, to say nothing of the outstanding taxonomic efforts that motivate her theory by making the case that none of the currently available theories of emotions singles out a scientifically fruitful domain of phenomena.

Moors (2022) makes a compelling case that emotions are not disturbances on the way to goal-pursuit, but manifestations of the ability to pursue goals. This is the insight that also powers my *Motivational Theory of Emotions* (MTE), according to which emotions are motivational structures producing action tendencies with control precedence when prompted by suitable stimuli (Frijda, 1986; Scarantino, 2014).

The difference is that I put stimulus-driven processes at the heart of emotions, and Moors does not. In a *stimulus-driven process*, a stimulus S activates an innate or learned stimulus-response link S-R, where R is an action tendency or intention:

> *Stimulus-driven process*: Stimulus S→ [S–R link] →action tendency/intention R is formed in a pre-determined fashion

MTE identifies emotions with behavioral systems pairing stimuli and action tendencies endowed with control precedence via a stimulus-driven process, where the tendency is to achieve a *goal* described at a high level of abstraction. For example, fear* is the behavioral system pairing appraised dangers with the prioritized tendency to avoid, anger* is the behavioral system pairing appraised goal-blockages with the prioritized tendency to aggress, and so on.

Corresponding author: Andrea Scarantino, Department of Philosophy, Georgia State University, 34 Peachtree Street, Atlanta, GA 30302-4089, USA.
Email: ascarantino@gsu.edu

These behavioral systems do not require dedicated neural circuitry and are not co-extensive with folk psychological emotion categories, because there are things we call fear and anger in English which differ from the behavioral systems of fear* and anger*. MTE is not trying to identify what ordinary people mean when they use the terms "fear" and "anger," but rather what is a "good thing to mean" by fear* and anger* for theoretical purposes. This process follows the guidelines laid out by Carnap (1962) in his theory of *explication*, understood as a process of conceptual revision of folk psychological categories into prescriptively defined categories similar to, but more fruitful than, the starting categories to be explicated.

I drop * in the rest of my commentary for simplicity of reference, but readers should keep in mind that the terms fear and anger will be used from here on as labels for prescriptively defined explicative categories (Carnap, 1962, Scarantino, 2012, Widen & Russell, 2010).

In a *goal-directed process*, the action tendency or intention R* results from a decision process which selects between different response options the one that maximizes expected value:

> *Goal-directed process*: S→ [Stimulus S: computation of R1's expected value Ov] [Stimulus S: computation of R2's expected value Ov]… → [R* is represented as having the highest expected value] →action tendency/intention R* is formed to maximize expected value

Moors states that "emotions are goal-directed cycles" (2022, p. 225) which get started with the detection of a discrepancy between a stimulus and a high value goal, are followed by the selection of an action to undo the discrepancy, and end with the comparison of the selected action's outcome with the initial stimulus, with the cycle continuing until the discrepancy is eliminated. Various phases of the cycle produce positive or negative feelings, which account for the valence of emotions.

For example, anger is on Moors' view not a behavioral system pairing appraised goal-blockages with the tendency to aggress, but a goal-directed cycle generated by detecting an unpleasant discrepancy between a stimulus and the high value goal of being respected, and involving the selection of an action to undo the discrepancy based on cost-benefit analysis. Similarly, fear is not a behavioral system pairing appraised dangers with the tendency to avoid, but a goal-directed cycle generated by detecting an unpleasant discrepancy between a stimulus and the high value goal of being safe, and involving the selection of an action to undo the discrepancy based on cost-benefit analysis.

Why does Moors think that the action tendencies typical of emotions are produced by a goal-directed rather than a stimulus-driven process? Her main argument is that stimulus-driven and goal-directed processes can both operate automatically, contrary to popular opinion, but the competition between them is generally "won by the goal-directed process because this process combines more benefits (automaticity and adaptiveness) than stimulus-driven processes do (only automaticity), and the system should prioritize the process with the most benefits" (Moors, 2022, p. 68).

I find this argument singularly unpersuasive. By *adaptiveness*, Moors means *practical rationality*, which is the ability of an emotion to satisfy the emoter's goals. In turn, practical rationality has two sides by Moors' lights: *practical rationality in the process sense*, which amounts to being a process of behavior causation that considers the value of action options and so has the potential to satisfy the emoter's goals, and *practical rationality in the outcome sense*, which amounts to being a process of behavior causation that actually satisfies the emoter's goals.

The fact that goal-directed processes have, and stimulus-driven processes lack, practical rationality in the process sense is not sufficient to conclude that they deliver the "most benefits." The benefits in terms of adaptiveness that lead a process of behavior causation to win over another have to do with outcome-rationality, namely with their comparative capacities to actually satisfy the emoter's goals. A behavior-causing process which does not consider the value of options would certainly win over another process that does, if it systematically led to higher satisfaction of the agent's goals.

The point is that we have no reason to think that stimulus-driven processes by default lead to lower satisfaction of the agent's goals compared to goal-directed processes. Take the goal of safety from mortal threats. It is hard to conceive of a better strategy to deal with mortal threats than one which begins with a pre-determined tendency to avoid them. Or take the goal of pursuing one's goals without obstruction. It is hard to conceive of a better strategy for dealing with goal-obstruction than one which starts from a pre-determined aggressive tendency aimed at removing the cause of the obstruction.

This is true because when high value goals like being safe or being respected are threatened, the operating conditions for the behavior-causing process tend to become poor—the time to act is limited, attentional resources become scarcer, cognitive processing capacities are reduced, access to information is curtailed and so on. Stimulus-driven processes have no trouble functioning in such circumstances, whereas it is an entirely open question whether goal-directed processes are similarly robust in poor operating conditions.

The second reason why I reject the view that goal-directed processes will crowd out stimulus-driven ones because of their superior adaptiveness is that we are presented with a false choice, namely the one between purely goal-oriented processes and purely stimulus-driven processes.

MTE's core proposal is that stimulus-driven processes and goal-directed processes cooperate rather than compete in the control of action: right after an action tendency has been generated via a stimulus-driven process, a goal-directed process of decision-making steps in. This is because action

tendencies allow for countless actions aligned with the tendency, as well as inhibition. For example, a goal-directed process must determine if and how the stimulus-driven aggressive tendency associated with anger is to be manifested, which may lead to diverse actions like yelling, spitting, kicking, shooting, sulking, slamming a door, calling the police, or to suppressing the tendency entirely.

A more relevant comparison in terms of outcome-rationality should pit behavior-causing processes that are either purely goal-directed or purely stimulus-driven with hybrid behavior-causing processes that combine a stimulus-driven phase (generation of the tendency) with a goal-directed phase (regulation of the tendency within milliseconds of its generation). It seems likely that hybrid processes will produce "more benefits" with respect to automaticity and adaptiveness than purely goal-directed ones, as they are robust in both ample and poor operating conditions and just as flexible as purely goal-directed processes once the action tendency is activated. For these reasons, I reject the claim that goal-directed processes are the default action tendency formation mechanism for emotions.

The troubles for the goal-directed theory do not end here. The identity claim at the heart of the theory—"emotions are goal-directed cycles"—is false. Suppose someone insults me, I detect a discrepancy with the goal of being respected, and select a mental or physical action to undo the discrepancy. Moors notes that there are three ways this goal-directed cycle could go, and she does not clarify which of the three ways instantiates the emotion, although the implication is that all three may instantiate it (see e.g., footnote 82 in Moors, 2022, p. 215, where she compares fainting out of fear to a form of immunization). First, I can drop the goal of being respected prior to having produced any aggressive tendency, and thereby remove the discrepancy (*accommodation*). Second, I can reinterpret the insult as having been said in jest and thereby dissolve the discrepancy prior to having developed any aggressive tendency (*immunization*).

It is quite clear that accommodation and immunization cycles do not instantiate anger, because the prototypical components associated with anger are missing in both cases. If I detect a discrepancy between being insulted and the goal of being respected, this can feel unpleasant, but this judgment + feeling combo is not an anger episode unless, at a bare minimum, I continue wanting to be respected (i.e., I do not accommodate) and I do not reinterpret the insult as a non-insult (i.e., I do not immunize).

The third way the goal-directed cycle could go is for Moors *assimilation*, which involves the formation of an action tendency to undo the discrepancy with the goal of being respected. When Moors says that emotions are goal-directed cycles, I understand her to have in mind first and foremost goal-directed cycles that involve *assimilation*, because these are the cases in which a physical action tendency is formed. Could anger be an *assimilation cycle* then*?* There are several reasons why the answer is a resounding no.

The first is that anger cannot possibly be instantiated if the "wrong" action tendency is selected. Suppose my cost-benefit analysis leads me to select a value maximizing action tendency which is *not* the manifestation of an aggressive tendency, like choosing to apologize to the person who insulted me assuming, rightly or wrongly, that this will lead them to apologize in return. The prototypical features of anger would in this case be sorely missing.

Note that this is different from *inhibiting* an aggressive tendency, with its distinctive suite of somatic and motoric underpinnings, because one considers it strategically better to apologize instead, a possible outcome on the hybrid MTE model. In Moors' model, no aggressive tendency is formed prior to the cost-benefit analysis, so if the agent ends up selecting apologizing as the value maximizing option, the agent only forms an apologizing tendency. But responding to an insult by forming a tendency to apologize to the person who insulted you without experiencing so much as a hint of aggression is not anger.

The second, and more damaging, problem for Moors' theory is that even selecting the "right" action tendency is not enough for instantiating anger. This is because there are non-emotional ways of forming action tendencies to undo discrepancies with high value goals. If my boss insults me and I detect a discrepancy with the goal of being respected, I can select the action of reporting him to HR with the aim of eliminating the discrepancy, but do so without any anger. Similarly, if a professional lion tamer detects a discrepancy between a behavior of the lion and the goal of safety, he may form the tendency to leave the cage while experiencing no fear.

It is of course also possible that I report my boss to HR *angrily*, or that the lion tamer exits the cage *fearfully*, but the point is that the distinction between emotional and non-emotional ways of dealing with high value goals cannot be accounted for by Moors' theory. One may retort that if the goals at stake were *truly* high value, emotions would inevitably come about, but this is nothing more than an *ad hoc* stipulation.

The implication of my discussion so far is not simply that there are goal-directed cycles which are not emotions, a point on which Moors and I agree, but that having an emotion is different in kind from engaging in a goal-directed process originated by detecting a discrepancy with a high value goal, even when that process leads to the formation of a physical action tendency (the assimilation case).

The key problem is this: although emotions are ways to deal with high value goals by generating action tendencies, not all ways to generate action tendencies to deal with high value goals are emotions. This is largely because not all action tendencies produced to deal with high value goals have *control precedence*, which MTE considers to be the "mark of the emotional" (Frijda, 1986; Scarantino, 2014).

Control precedence involves bodily and mental preparation, goal prioritization, focused attention, and a roster of distinctive biases in accessing information, recruiting memories, drawing inferences, evaluating evidence, assigning probabilities, assessing risks, acquiring inertial properties, and so on (Scarantino, 2014). Trying to eliminate discrepancies with high value goals may or may not lead to forming action tendencies with control precedence.

Could Moors simply react to the charge that emotions are not goal-directed cycles with high value goals by advocating that, since folk psychological emotion concepts are too heterogeneous to allow for scientific extrapolation, we don't really need to worry about accounting for emotions in goal-directed terms? Something along these lines seems to underlie her remarks that "even if the goal-directed theory covers all typical emotional components, it remains agnostic about which of these components should be part of a constitutive explanation of emotion. The theory has a skeptical outlook and merely tries to explain the phenomena that people call emotions" (2022, pp. 217–218).

But it is hard to see how one can reconcile agnosticism about what components of the goal-directed cycle instantiate emotions with the ambition of explaining the phenomena people call emotions. All forms of eliminativism about emotion face a basic challenge: showing us that a theory which replaces emotions with something else is not worse off explanatorily than a theory which retains emotions. It is not enough to argue that folk emotion concepts are defective—an eliminativist must also make sense of the phenomena emotion concepts appear to explain.

This explanatory task is not met by the current version of the goal-directed theory, where the high value of goals is the only resource available to explain the difference between emotional and non-emotional behaviors. I offer two examples of the acrobatics Moors (2022) must engage in to try and make sense of emotional phenomena in terms of cost-benefit analysis: decisional paralysis and recalcitrance.

Sometimes, emotions are so intense that they short-circuit our capacity to make decisions. When a loaded gun is pointed at your face by a mugger, you may experience a type of fear which paralyzes you. You fail to listen to instructions, do not relinquish your wallet when asked, and end up getting shot. MTE proposes that you form a stimulus-driven avoidance tendency upon detecting the gun, but are unable to determine if and how to manifest it—your decision-making capacity is thwarted by the intensity of fear itself.

Trying to explain this episode in terms of a goal-directed cycle involves positing a cost-benefit analysis of the available action options prompted by the detection of a discrepancy between the gun and safety, followed by the representation that no action has a higher expected utility than doing nothing, followed by the value-maximizing choice of doing nothing (Moors & Fischer, 2019). This gravely misconstrues what goes on in emotionally induced decisional paralysis, because choosing to do nothing is an action, whereas being paralyzed by fear is something that happens to an emoter. A person paralyzed by fear is *not* someone who lacks any tendency to avoid what they fear and who deliberately chooses to do nothing because of lack of options with a sufficiently high expected value.[1]

Emotions can also be recalcitrant, that is, in conflict with our beliefs, which is another one of the ways in which they can become irrational. I may believe that a spider caged in a transparent box at the zoo is not dangerous, and yet flee it despite my best efforts not to do so. MTE explains this case as follows: I first form a stimulus-driven avoidance tendency upon detecting the spider via an appraisal system which is partially cognitively impenetrable, then I perform a cost-benefit analysis concluding I should stand still to contemplate the spider, and finally I fail to do what I deliberated to do and flee out of fear, because the fear-induced avoidance tendency is stronger than the action tendency to contemplate the spider, and cannot be suppressed.

My belief that the spider is not dangerous fails to change my fear appraisal, creating an *epistemic inconsistency* between my emotional construal and my belief. As a result, fear motivates me to avoid the spider, and I end up doing so despite having other goals in conflict with it, like the goal of not coming across as a coward to my friends, the goal of completing the zoo visit and so on. For MTE, recalcitrance provides compelling evidence that fear is a distinctive behavioral system with its own set of activating triggers.

Moors (2022, pp. 229–230) tries to account for recalcitrance in terms of a conflict between goal-directed cycles driven by different goals: bodily integrity, epistemic consistency,[2] and impression management, presumably with respect to the people around. If the goal of bodily integrity has higher value than epistemic consistency and impression management, the agent will choose to flee in order to undo the activated discrepancy between the spider and the goal of bodily integrity. The problem is that this seriously misrepresents what goes on in cases of recalcitrance, because the agent is *not deliberately choosing* to give up on epistemic consistency and impression management to preserve bodily integrity.

First, the agent does not voluntarily relinquish epistemic consistency to achieve some other goal—this is the whole point of recalcitrance. If it was up to her, the agent would choose to have her emotional construal aligned with her belief, that is, no epistemic inconsistency at all, but this is something she cannot achieve despite her best efforts. Second, if the agent believes that there is no discrepancy, the agent also believes that bodily integrity is already secured. Moors notes that "[i]n the case of the spider, a discrepancy with the goal for safety may be activated even if the person does not believe that…she is in danger," but this does not explain why the agent "chooses to be on the safe side" by fleeing (2022, p. 230).

Someone can fear a spider caged in a transparent box *recalcitrantly* only if they believe their bodily integrity has

been secured. If they are not entirely sure, or if they are merely imagining threats to bodily integrity without having settled on whether such threats are instantiated, recalcitrance (as commonly defined) vanishes, because there is no epistemic conflict between fearing a caged spider and believing there is a chance such spider may be a threat to bodily integrity or just imagining the spider becoming a threat.

But if the agent is convinced that bodily integrity is already secured, fleeing the spider cannot be the option with the highest expected value, because the agent believes fleeing will *prevent* the achievement of two goals (epistemic consistency and impression management) without delivering *any* improvement towards the goal of bodily integrity. If anything, an agent governed by a goal-directed process would select the strategy of *immunization* once the goal of bodily integrity is activated, namely the reinterpretation of the spider as compatible with the goal of bodily integrity. And yet, the recalcitrant agent cannot bring herself to select the immunization strategy either. What recalcitrance calls for, explanatorily speaking, is a stimulus-driven process that can hijack behavior by directly generating an avoidance tendency so intense that regulation based on expected values is interfered with.[3]

This is just the tip of a large iceberg, as the differences between acting emotionally and acting in the pursuit of high value goals by maximizing expected values can be multiplied at will. The upshot is that the eliminativist project of replacing emotions with goal-directed cycles radically reduces our ability to understand a variety of behavioral effects of emotions.

My hope is that Moors will consider giving stimulus-driven processes a more substantial role to play in the explanation of emotional phenomena, well beyond the two cases of stimulus-driven processes she explicitly considers in Moors (2022). One is the case of agents facing equally attractive options (Buridan's case), where rational decision-making would stall if stimulus-driven bias did not step in. The other is the case of agents acting without activating goal-directed cycles, as one does when automatically pressing an elevator button to a certain familiar floor without any goal-directed process commanding the pressing of a different button, in which case stimulus-response links operate unopposed.

Neither variety of stimulus-driven process is directly applicable to decisional paralysis or recalcitrance, but, as we have seen, goal-directed processes alone cannot make sense of them. This reveals a need for granting stimulus-driven processes a more central role in the emotional realm. The most promising explanation of why emotions interfere with, or are isolated from, other parts of the rational mind is that they rely on a proprietary action tendency formation mechanism which can occasionally interfere with cost-benefit analysis.

Expanding the role of stimulus-driven processes would undercut Moors' eliminativist attempt to make sense of emotions without relying on any emotion-specific mechanisms. But it would not undermine the key role goal-directed processes play in emotional phenomena, sometimes as *substitutes* for stimulus-driven processes, sometimes as *partners* of stimulus-driven processes, and in yet other cases as *competitors* of stimulus-driven processes for the control of behavior.

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 received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Andrea Scarantino https://orcid.org/0000-0003-1396-8128

Notes

1. Kilic (2025) discusses several other ways in which Moors could try to account for cases of decisional paralysis in goal-directed terms, and finds them all unpersuasive.
2. Moors (2022) understands epistemic inconsistency as a conflict between one's belief and "one's other knowledge" (229), but this is incorrect, because the recalcitrant agent believes, consistently with his other knowledge, that spiders are not dangerous. The actual conflict at stake in cases of epistemic inconsistency is the one between one's belief and one's emotional construal.
3. Pineda-Oliva (2025) considers other ways in which Moors could try to account for recalcitrance, but finds them all inadequate.

References

Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). University of Chicago Press.

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

Kılıç, O. (2025). *Practical irrationality mystified* [MA thesis]. Georgia State University. https://scholarworks.gsu.edu/items/6cf7466d-3e89-45be-b8e9-8699f8e66f0f

Moors, A. (2022). *Demystifying emotions*. Cambridge University Press.

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

Pineda-Oliva, D. (2025). Some reflections on the goal-directed theory of emotion. *Acta Analytica*. https://doi.org/10.1007/s12136-025-00637-3

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

Scarantino, A. (2014). The motivational theory of emotions. In D. Jacobson & J. D'Arms (Eds.), *Moral psychology, and human agency* (pp. 156–185). Cambridge University Press.

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