To cite this article:

Moors, A. (2025). When unpacking the black box of motivation invites three forms of reductionism. *Behavioral and Brain Sciences*, *48*, Article e37. https://doi.org/10.1017/S0140525X24000426

Commentary to Murrayama & Jach: A critique of motivation constructs to explain higher-order behavior: We should unpack the black box.

1. Agnes Moors

2. word count: abstract: 53; main text: 999; references: 137; entire text: 1296

3. Title: When unpacking the black box of motivation invites three forms of reductionism

4. Agnes Moors

5. KU Leuven – University of Leuven

6. Tiensestraat 102, 3000 Leuven, Belgium

7. Tel. +3216373279

8. agnes.moors@kuleuven.be.

9. http://ppw.kuleuven.be/okp/people/Agnes_Moors/

10. Abstract

In their proposal for unpacking the black box of motivation, Murrayama and Jach (2024) propose three types of reductions: from high-level to low-level motivational constructs, from motivation to cognition, and from contentful to contentless explanations. Although these reductions come with the promise of parsimony, they carry the risk of losing vital explanatory power.

11. Main text

The motivation literature typically distinguishes between process theories and content theories. Process theories seek to provide a mechanistic explanation of behavior, specifying a mechanism between input (stimulus) and output (behavior) that includes motivation constructs such as goals. Content theories address questions about what our fundamental goals

are (e.g., autonomy, competence, and connectedness in self-determination theory; Ryan & Deci, 2017), or they apply existing process theories to a specific type of goal (e.g., achievement, affiliation, power; Murray, 1938). Unlike content theories, existing process theories have already taken steps toward opening the black box that sits between observable stimulus input and behavioral output. The critique of Murayama and Jach (2024) therefore seems to be more directed toward content theories than toward process theories. Alternatively, their critique could be directed toward process theories, suggesting that these theories do not unpack their mechanisms in sufficient detail, that is, at a low enough level of analysis.

The authors go on to propose a way to do this unpacking. Despite cautioning that their exercise is not a reductionist attempt, they appear to promote three types of reductions. I will argue how each of these types poses risks for throwing away the baby with the bathwater.

A first type of reduction is a shift from high-level (abstract) motivational constructs to low-level (concrete) motivational constructs. For instance, they propose replacing the goal for "competence" with the goal for "information seeking". The ideal seems to be to shift to motivational constructs that are as close as possible to the behavior itself. Thus, they argue that information-seeking behavior is caused by the goal to have information rather than by the goal for competence. They further propose that such a low-level goal is caused by a gap in this goal. Thus, the goal to have information is caused by uncertainty, that is, a gap between a current and desired level of information.

The ultimate step in this shift toward low-level constructs is the shift from more extrinsic to more intrinsic motivation. People seek information because they value the act of information seeking (or at least the immediate outcome of this act: information) rather than because this

information is instrumental for reaching other goals. Such a shift seems to downplay the role of extrinsic motivation, however. There may certainly be cases in which people act for the sake of it, but much of our behavior, including information seeking, is done to reach other goals. The explanatory power of existing process theories resides in the fact that they propose a goal hierarchy in which behavioral goals that are the proximal causes of behavior can be considered as subordinate goals that have a high expectancy for reaching other, superordinate goals.

This feature is preserved in several existing process theories, including in my own goal-directed theory of behavior causation (Moors, 2022; Moors et al., 2017). Here a gap between a high-level goal and a current state activates the goal to reduce this gap. This can be done by different types of behaviors (i.e., assimilation), but also by adjusting the high-level goal (i.e., accommodation) or by changing the interpretation of the current state (i.e., immunization). To illustrate, the gap between the current state and the high-level goal to become popular activates the goal to reduce this gap. One way to reduce this gap is via behavior (e.g., wearing nice cloths, making jokes; i.e., assimilation), another is to give up the goal to become popular (i.e., accommodation), and still another is to reinterpret the current state as one in which you are already popular (i.e., immunization).

It might be argued that the high-level goals proposed in content theories such as the goals for autonomy and competence are not well chosen, but then the solution would be to come up with better ones, not to do away with high-level goals altogether. An alternative solution would be to view the goals for autonomy and competence as meta-goals that are at the service of, or assist in, the attainment of other, low-level goals. In this vein, the goal for autonomy can

be considered as the goal to be allowed to *choose* one's own (low-level) goals and the goal for competence or control can be considered as the goal to *achieve* these (low-level) goals.

A second type of reduction that Murrayama and Jach (2024) seem to promote consists in a shift from motivation to cognition. The authors admit that in their computational process model, there are still rewards, which are representations of valued outcomes and hence motivational constructs. However, once the unpacking of the black box has arrived at its most concrete level of motivation, the authors argue that it makes little sense to continue calling this motivation. This reveals that the ideal to which they aspire is to ultimately reduce motivation to cognition. This is reminiscent of the attempt of predictive processing theory to reduce the explanation of behavior (and other phenomena such as perception and affect) to the confirmation and disconfirmation of expectations (Clark, 2013; Friston, 2009).

A third type of reduction that the authors seem to promote consists in gradually explaining away content or semantics. In standard mechanistic explanations in psychology, mechanisms between stimulus input and behavioral output are composed of representations with a content and a format (i.e., the structural parts) and operations acting on these representations (i.e., the activities or working parts) (Bechtel, 2008). The authors' ideal seems to be to reduce mechanistic explanations that consist of both contentful representations and operations to explanations that consist only of operations.

In conclusion, even if existing process theories of motivation have already made progress in unpacking the black box, it may be argued that there is always room for further unpacking at lower levels of analysis. Whether this shift to lower levels of analysis should include the rejection of high-level goals, the reduction of motivation to cognition, and the evolution

toward content-less mechanistic explanations is open to debate, as is the level of analysis that will prove to be most fruitful for predicting and regulating behavior (Karoly, 1999).

12. Acknowledgment statement: none

13. Competing interests: none

14. Funding statement: Preparation of this article was supported by Grant C14/23/062 of the Research Fund of KU Leuven.

15. Reference list:

Bechtel, W. (2008). Mental mechanisms: Philosophical perspectives on cognitive neuroscience. Routledge. https://doi.org/10.2478/disp-2008-0012

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204. https://doi.org/10.1017/s0140525x12000477

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, *13*(7), 293-301. https://doi.org/10.1016/j.tics.2009.04.005

Karoly, P. (1999). A goal systems–self-regulatory perspective on personality, psychopathology, and change. *Review of general psychology*, *3*(4), 264-291. https://doi.org/10.1037//1089-2680.3.4.264

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

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*, 310-318. https://doi.org/10.1177/1754073916669595

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* Guilford. https://doi.org/10.1521/978.14625/28806