

## Personal and Subpersonal Explanation: The Interface Problem Revisited

977 words without references

### Extended abstract

In the philosophy of cognitive science, it is customary to talk about levels of explanation. There are different conceptualizations of how such talk is to be understood. While contemporary work often frames levels in mechanistic terms (e.g. Francken et al., 2022; Piccinini, 2020), philosophers have also drawn a more general distinction between personal and subpersonal levels of explanation (e.g. Drayson, 2024; Jurjako & Wilkinson, 2025; Musholt, 2018). This paper focuses on that distinction.

A central unresolved issue concerns how the personal and subpersonal levels of explanation are related. Unlike debates about levels within mechanistic explanation, this relation appears to involve different kinds of explanation altogether. This issue is often described as the interface problem (Bermúdez, 2005; Drayson, 2024). However, there is little agreement even on how the interface problem itself should be understood. José Bermúdez (2005) characterizes the interface problem as reflecting competing ways of understanding the personal/subpersonal distinction itself, distinguishing autonomist, functionalist, and co-evolutionary models. In contrast, Zoe Drayson (2012, 2014, 2024) argues that only a functionalist understanding of the distinction is viable and treats the interface problem as specific to that framework.

I argue against Drayson's view and defend a pluralist understanding of the interface problem, according to which there are multiple legitimate ways of relating personal and subpersonal explanations. I support this claim by revisiting Daniel Dennett's (1969) original introduction of the personal/subpersonal distinction.

The personal/subpersonal distinction is often understood as involving two different types of explanation. According to autonomism, the personal and subpersonal levels are understood as independent because they involve different types of explanation governed by different kinds of constraints (e.g. Davidson, 2001; Hornsby, 1997; McDowell, 1994). Personal explanations make people's behavior intelligible in terms of folk-psychological constructs such as beliefs, desires, and intentions, which are intentional and governed by rational constraints. Subpersonal explanations, by contrast, are often characterized as non-intentional and non-rational, and as proceeding by identifying mechanistic components—such as parts of a neural system—that enable psychological capacities. This contrast motivates the autonomist resistance to applying psychological predicates to components of persons (Bennett & Hacker, 2022).

In contrast, Drayson understands the personal/subpersonal distinction primarily in terms of part-whole relations. On her view, personal explanations ascribe psychological predicates to whole individuals, whereas subpersonal explanations ascribe psychological predicates to parts of individuals. Drayson (2012) argues that this way of drawing the distinction traces back to Dennett's original motivation for introducing it, namely, to legitimize a functionalist/homuncular methodology in cognitive science. On this approach, cognitive capacities are decomposed into subcapacities modelled as task-performing "little agents". Consequently, while psychological attributes such as beliefs and desires are ascribed to whole agents at the personal level to explain behavior, similar psychological attributes—such as representations, preferences, or predictions—may also be ascribed to neural subsystems, as is common practice in contemporary cognitive neuroscience (e.g. Pezzulo et al., 2018; Smith et al., 2022).

On this basis, Drayson (2012) argues that autonomist and functionalist approaches do not even share a common understanding of the personal/subpersonal distinction, and thus should not

be seen as offering competing solutions to the interface problem. According to her, only functionalist approaches genuinely address the personal/subpersonal distinction, while autonomist views are better understood in terms of Wilfrid Sellars's space of reasons/space of causes distinction (Drayson, 2014).

Against Drayson, I argue that Dennett's (1969) original formulation of the personal/subpersonal distinction supports both autonomist and non-autonomist readings. I substantiate this claim through an analysis of Dennett's treatment of pain, reasons, and consciousness.

Regarding pain, Dennett adopts the autonomist approach. Dennett claims that pain figures only in autonomous personal-level explanations. Pain is, of course, correlated with afferent and efferent neural systems that link certain stimuli to characteristic behavioral responses. However, when we move to this subpersonal level of explanation, Dennett maintains that "in a very real sense we abandon the subject matter of pains as well" (Dennett, 1969, pp. 93–94).

Regarding reason-based explanations, Dennett is a functionalist. About the relation between reasons and causes, he claims that

"(...) not only is it the case that when I do something for a reason, what I do is caused, but what makes a reason my *real* reason for doing something is that the events of information processing which cause what I do have among them an event with the content of my real reason (...). (Dennett, 1969, p. 163, emphasis in the original)

Here, Dennett appears to come close to a language-of-thought–style view, on which an agent's reasons at the personal level are treated as encoded in subpersonal information-processing events.

In other contexts, Dennett endorses a revisionary view according to which subpersonal considerations can shape the meaning of personal-level terms. This is clear in Dennett's discussion of consciousness, where he distinguishes between awareness<sub>1</sub>, which involves the reportability of experience via a subpersonal linguistic module, and awareness<sub>2</sub>, which does not (Dennett, 1969, p. 118). Dennett is explicit that the introduction of the terms awareness<sub>1</sub> and awareness<sub>2</sub> is intended to

“bridge the gap between the personal and the subpersonal. The ordinary personal-level term ‘aware’ is being replaced by two terms that still take persons (or whole systems) as subjects, but have subpersonal-criteria” (Dennett, 1969, p. 119).

Here, subpersonal considerations are used to supply and revise criteria for the application of a personal-level term such as “awareness”.

I argue that these considerations undermine Drayson's view in two ways. First, Dennett's original introduction of the distinction does not commit him to a fixed account of the relation between personal and subpersonal explanations: personal-level constructs may be autonomous; map onto subpersonal processes; or be revised in light of them. Second, this flexibility supports a shared, neutral understanding of the personal/subpersonal distinction, within which autonomist and non-autonomist views should be seen as competing solutions to the interface problem rather than as reflecting incompatible conceptions of the distinction.

## References

- Bennett, M. R., & Hacker, P. M. S. (2022). *Philosophical foundations of neuroscience* (Second Edition). John Wiley & Sons.
- Bermúdez, J. L. (2005). *Philosophy of psychology: A contemporary introduction*. Routledge.

- Davidson, D. (2001). *Essays on actions and events*. Oxford University Press.  
<https://doi.org/10.1093/0199246270.001.0001>
- Dennett, D. C. (1969). *Content and consciousness*. Routledge.
- Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives*, 26(1), 1–18. <https://doi.org/10.1111/phpe.12014>
- Drayson, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass*, 9(5), 338–346.  
<https://doi.org/10.1111/phc3.12124>
- Drayson, Z. (2024). Personal/subpersonal distinction. In M. C. Frank, A. Majid, & A. Alsmith (Eds.), *Open Encyclopedia of Cognitive Science*. MIT Press.  
<https://doi.org/10.21428/e2759450.8593aeda>
- Francken, J. C., Slors, M., & Craver, C. F. (2022). Cognitive ontology and the search for neural mechanisms: Three foundational problems. *Synthese*, 200(5), 378.  
<https://doi.org/10.1007/s11229-022-03701-2>
- Hornsby, J. (1997). *Simple mindedness: In defense of naive naturalism in the philosophy of mind*. Harvard University Press.
- Jurjako, M., & Wilkinson, S. (2025). The personal and the subpersonal: Three desiderata and a pragmatist proposal. *Philosophical Studies*, 182, 2875–2904.  
<https://doi.org/10.1007/s11098-025-02381-1>
- McDowell, J. H. (1994). The content of perceptual experience. *The Philosophical Quarterly*, 44(175), 190. <https://doi.org/10.2307/2219740>
- Musholt, K. (2018). The personal and the subpersonal in the theory of mind debate. *Phenomenology and the Cognitive Sciences*, 17(2), 305–324.  
<https://doi.org/10.1007/s11097-017-9504-4>

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences*, 22(4), 294–306.

<https://doi.org/10.1016/j.tics.2018.01.009>

Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*. Oxford University Press.

Smith, R., Ramstead, M. J. D., & Kiefer, A. (2022). Active inference models do not contradict folk psychology. *Synthese*, 200(2), 81. <https://doi.org/10.1007/s11229-022-03480-w>