

Beyond Deliberation: How Memory Shapes Intentions

Consider an agent who forms an intention to pursue a particular career, judging the path well suited to her. Months pass. She never sits down to reconsider, and no decisive new information arrives. Yet when the plan next comes up it no longer moves her as it once did: remembered disappointments are more salient, the imagined future less appealing. What changed for her, and how does it bear on accounts that treat deliberation as the main driver of intention revision?

Leading theories of temporally extended agency offer only partial answers. Planning theory, in Bratman's tradition, treats intentions as conduct-controlling and reasoning-anchoring states, stable across time and revised through deliberation (Bratman, 1987; Holton, 2009). Valuationism presses a rival picture: action runs on value representations consulted at each decision, and intentions carry little explanatory weight (Sripada, 2025). The two seem to emphasize different timescales of action coordination. But neither fully explains the processes gradually affecting a standing commitment across the stretches when the agent is not deliberating.

I argue that this interval is not inert. Because intention must be stored in memory until the time to act arises, intentions are likely reworked offline by processes known to affect memory. Value-updating, through experience replay, modifies the expected value tied to the intention, and so its motivational pull (Liu et al., 2021). Affect-tagging modifies affective valence, altering both pull and accessibility. Semantization makes the content of memories more abstract and less episodically detailed over time (Aronowitz, 2025). None of these processes amounts to deliberation: each runs while the agent is at rest, accumulating in small increments and reaching no conclusion she reflectively avows. Two pathways should be distinguished: a change to stored representations count as a change *in* the intention when it alters the representations constituting the intention itself that are responsible for conduct-control or reasoning-anchoring; otherwise, it is a change *around* the intention – say, to the reasons on which the intention is based.

What determines when offline change reaches deliberation? I argue for a metacognitive controller that weighs the benefit of deliberating against its cost, engaging it only once certain signals – an affective marker, conflict, or shifts in value or content – cross a threshold. The agent registers the shift only when offline changes have been substantial enough to warrant engaging deliberation. This fits naturally with the machinery native to valuationism.

However, valuationism should also acknowledge the role of intentions in coordinating our activities over time in the pursuit of long-term goals and in reducing cognitive load. A standing intention filters: inconsistent options tend to drop out before their value is computed. A gate keyed purely to value cannot do this, since the excluded option may be one the agent would value highly. The filter also answers to consistency with our prior commitments. Representational transformation re-keys this filter: as the commitment abstracts from a particular job toward a career of a certain shape, a comparable job elsewhere, once ruled out, becomes admissible – a change in consistency, not just value.

The result is a hybrid account: planning theory's farsighted intentions and valuationism's nearsighted value representations and updates become parts of one system, with incremental transformations within memory affecting the representations guiding our actions between deliberative episodes.

References

- Aronowitz, S. (2025). Semanticization Challenges the Episodic–Semantic Distinction. *British Journal for the Philosophy of Science*, 76(4), 805–822.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.
- Holton, R. (2009). *Willing, Wanting, Waiting*. Oxford University Press.
- Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science*, 372(6544).
- Sripada, C. (2025). The valuationist model of human agent architecture. *Philosophical Psychology*, 1–30.