

University of Memphis

University of Memphis Digital Commons

CCRG Papers

Cognitive Computing Research Group

2021

Modeling long-term intentions and narratives in autonomous agents

C. Kronsted

Z. Neemeh

S. Kugele

S. Franklin

Follow this and additional works at: https://digitalcommons.memphis.edu/ccrg_papers

Recommended Citation

Kronsted, C., Neemeh, Z., Kugele, S., & Franklin, S. (2021). Modeling long-term intentions and narratives in autonomous agents. [10.1142/S2705078521500107](https://digitalcommons.memphis.edu/ccrg_papers/10.1142/S2705078521500107)

This Document is brought to you for free and open access by the Cognitive Computing Research Group at University of Memphis Digital Commons. It has been accepted for inclusion in CCRG Papers by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.

Modeling Long-Term Intentions and Narratives in Autonomous Agents

Christian Kronsted

*Department of Philosophy and Institute for Intelligent Systems, The University of University of Memphis,
3720 Alumni Ave Memphis, Tennessee, 3815, United States of America
Christian.k@memphis.edu*

Zachariah A. Neemeh

*Department of Philosophy and Institute for Intelligent Systems, The University of University of Memphis,
3720 Alumni Ave Memphis, Tennessee, 3815, United States of America*

zaneemeh@memphis.edu

Sean Kugele

*Computer Science Department and Institute for Intelligent Systems, The University of University of
Memphis, 3720 Alumni Ave Memphis, Tennessee, 3815, United States of America*

skugele@memphis.edu

Stan Franklin

*Computer Science Department and Institute for Intelligent Systems, The University of University of
Memphis, 3720 Alumni Ave Memphis, Tennessee, 3815, United States of America*

franklin@memphis.edu

Received (Day Month Year)

Revised (Day Month Year)

Across various fields it is argued that the self in part consists of an autobiographical self-narrative and that the self-narrative has an impact on agential behavior. Similarly, within action theory, it is claimed that the intentional structure of coherent long-term action is divided into a hierarchy of distal, proximal, and motor intentions. However, the concrete mechanisms for how narratives and distal intentions are generated and impact action is rarely fleshed out concretely. We here demonstrate how narratives and distal intentions can be generated within cognitive agents and how they can impact agential behavior over long time scales. We integrate narratives and distal intentions into the LIDA model, and demonstrate how they can guide agential action in a manner that is consistent with the Global Workspace Theory of consciousness. This paper serves both as an addition to the LIDA cognitive architecture and an elucidation of how narratives and distal intention emerge and play their role in cognition and action

Keywords: LIDA; Global Workspace Theory; Narrative; Action Theory; Intentions; Distal Intentions; Agency.

1. Introduction

What separates agential action from mere behavior? When examined over a short period of time, autonomous agents and non-autonomous agents often produce behaviors that are similar in execution and outcome. For example, we can imagine an autonomous agent and a non-autonomous agent that can both pick up objects and put them in a trash can. However, a key difference emerges between the two since the actions of autonomous agents stem from an *agenda*. In order to carry out their agenda, autonomous agents must have a control structure that perpetually helps them answer the question: what do I do next? [Franklin, 1995, p. 16]. While fairly simplistic agents can act in accordance with their agenda on a moment-to-moment basis, more complex agents such as human beings have agendas that include aims that can only be fulfilled over time: what in the action theory literature has been dubbed our distal intentions [Mylopoulos & Pacherie, 2018a; Pacherie, 2006, 2008]. Saving for a house or training to become a great poet requires ordering one's actions in a fashion that is aimed at the distal future. For complex agents, agential action is characterized by setting and pursuing distal intentions^a. This means that complex autonomous agency is characterized in part by the ability to construct, follow, and adhere to long-term goals.

While very simple autonomous agents might not need distal intentions to pursue their agenda, complicated agents such as human beings may need nested structures of intentions. If such structures were not in place, we would have to renegotiate our major life choices and aims on a daily or even moment-to-moment basis. Distal intentions are forward looking intentions that contain a self-commitment. It is precisely because they contain such self-commitment that distal intentions have inertia – they are not easily changed or broken [Bratman, 1987]. While it is true that agents sometimes set distal intentions without committing to them, genuine distal intentions do structure behavior. For example, a person's hedonistic love of Philly cheesesteaks would simply win every time they are hungry if they were not able to construct and adhere to intentions such as "I want to get a six-pack," or "I want to compete in the fall triathlon." To avoid hedonistic behavior, autonomous agency and coherent action require the setting and following of distal intentions. Importantly, deliberate action is not just about juggling values from moment-to-moment, but also about adhering to one's own long-term goals.

Furthermore, much research has shown that in human agents, autonomy, the self, and personal identity are intimately tied to a self-narrative. We humans continually create, update, and follow narratives about ourselves with which to make sense of the world and our own actions [Bruner, 1991, 2004; Dings, 2019; Hutto, 2008; MacIntyre, 2013; Ryan *et al.*, 2019; Schechtman, 1996, 2014]. Our self-narratives and our intentions exist in a multicausal relationship to guide and structure our lives and our actions. Without intentions

^a In this paper, we use the word intention as it is used in the philosophical literature on action theory. In common English, the word 'intention' implies commitment. As we will see, in the technical, philosophical sense, there is a difference between intentions and intentions that have been committed to.

and self-narratives, many instances of volitional decision-making would end in stifled indecision. For complex agents such as ourselves, there is a causal relationship between the stories we create about ourselves and our conduct in the world. The kinds of stories relevant to our discussion in this paper are causal stories, namely stories that order typically temporally-distant events into a causal order. Self-narratives not only provide a sense of who the agent is, but do so by organizing important life events into a causal order.

While action theory can be traced back to the very early beginnings of recorded philosophy, very few action philosophers have said much about how the structure of intentional action is immersed into the structure of functional consciousness. Authors such as Anscombe [1979], Davidson [1985], Goldman [1970], Mele [1992], and Bratman [1987, 2004] have provided cognitive accounts of how action is structured and carried out, but these ignore the functional roles of consciousness and the cognitive cycle. How does the intention-action framework fit in with the cognitive structure of consciousness?

Further, while various disciplines are rife with discussion regarding the role of self-narratives in identity and cognition, these debates tend to happen at a fairly abstract level of explanation. In fact, very few have provided concrete explanations for how self-narratives are generated, stored, utilized, and impact agential behavior. Thus, this paper is not only an addition to the LIDA model, it also fills a gap by providing a very concrete model for how narratives and intentions support cognition.

In this paper, we apply the three tiered structure of intention, distal, proximal, and motor intention [Mylopoulos & Pacherie, 2018b; Pacherie, 2006, 2008] and narrative identity [Schechtman, 2014] to the study of functional consciousness through the LIDA model [Franklin *et al.*, 2016]. We show how the LIDA model can provide a systems-level account of narrative identity and the intentions that structure action through the cognitive cycle. While we will touch upon proximal and motor intentions, most of this paper will focus on narratives and distal intentions. Further, it is important to note for those readers not familiar with the philosophical debates on “intention”, that the notions of “intentions” and “intending” are much broader than how they are used in quotidian language. Intentional action refers to the embodied and cognitive processes that make agential action structured as opposed to random or reactive movements. So far, the LIDA model literature has mostly addressed processes occurring over a single cognitive cycle. This paper will provide an addition to the LIDA model that begins to sketch how LIDA agents act coherently over multiple cognitive cycles.

The LIDA model is an embodied approach to the system-level modeling of cognition. The LIDA model should, in principle, be able to satisfy both more radical embodied approaches to cognition as well as more traditional cognitivist approaches. The point here is that the LIDA model does not make any commitments to how the underlying mechanism of phenomenal consciousness is produced in brains, computers, or elsewhere.^b While the LIDA model aims to be compatible with neuroscience, it makes no claims about its physical implementation in brains. That being said, LIDA agents are still conceptualized

^b See Franklin *et al.* [2013] for a list of LIDA commitments.

as embodied, embedded in their environments, and always in the business of figuring out what to do next. To answer the question “what do I do next”, intentions, distal intentions, and self-narratives are all connected through the cognitive cycle, to produce coherent agential actions over long time scales.

In the past, the LIDA model has used the terminology of ‘options’ and ‘goals’ to capture the intentional nature of action. In this paper, to remain consistent with the literature on agential action in cognitive science and action theory, we use the term ‘intention.’ For the reader already familiar with the LIDA model, an intention is a goal. Additionally, and importantly for this paper, a *distal intention* is an agent’s goal that cannot be immediately fulfilled. Generally speaking, we use the words ‘goal’ and ‘intention’ synonymously.

2. What is LIDA?

The LIDA model is a systems-level cognitive architecture developed to model minds, whether human, animal, or artificial. LIDA is composed of multiple, mostly asynchronous modules, but makes no commitments to the modularity of underlying brain structures. Rather, LIDA explains, at the systems-level (or what some philosophers call ‘mid-level’ explanation), the structures at work in cognition and consciousness, and their interactions through each cognitive cycle.

One of the core notions is that “minds” are control structures for autonomous agents [Franklin, 1995; Franklin & Graesser, 1997]. Minds can be implemented in biological bodies, artificial bodies, or computers. The form of body does not matter, as long as the mind functions as a control structure allowing the agent to pursue its agenda. Minds are always in the business of answering the perpetual question “what should I do next?” The answer to such a question highly depends on the agent’s constitution – e.g., is it embodied, what kind of environment is it embedded within, and what is its overarching agenda? The LIDA model can account for some agents whether they are embodied or non-embodied agents (for example, biological agents such as humans, animals or robots, versus software agents that do not have a body in the conventional sense). For an account of how software agents can be embodied see [Franklin, 1997]. Regardless, as long as we are dealing with an autonomous agent, the mind of the agent cannot be divorced from its agenda.

Furthermore, the LIDA model is committed to action and cognition being smoothly integrated. Cognition happens through rolling, overlapping cycles in which the agent samples and acts on its environment. While most modules operate asynchronously, the broadcasts coming from the Global Workspace unfold in serial order to guarantee the coherent unfolding of *consciousness*. Additionally, Action Selection also operates synchronously and guarantees the coherent choice of *actions*. The sequential order of broadcasts aids the agent in selecting and executing actions that fit with the causal unfolding of events outside the agent. While most modules continually carry out their own,

independent functions, the sequential order of the broadcasts ensures that agents will act cohesively.

Importantly, the LIDA model is based on the Global Workspace Theory of consciousness [Baars, 1988, 1997, 2019]. Consciousness is included in ongoing cognitive cycles, in which each cycle is one cognitive moment or atom. Cognition consist of cycles of recurring events, wherein the rapid nature of the cycles create the fluidity of conscious experience (at an estimated ~10Hz in humans) [Madl *et al.*, 2011]. Each cycle is a very rapid cognitive moment in which the agent perceives the environment, attends to that environment, and finally chooses an action (for humans, all within roughly 200-500 milliseconds). The cognitive cycle is divided into three overarching phases: an understanding phase, an attending phase, and an action selection/learning phase. In LIDA learning (updating of the modules) can take place across most modules with each and every cognitive cycle [Kugele & Franklin, forthcoming].

At the beginning of each cognitive cycle, the agent's Current Situational Model (CSM) in the pre-conscious Workspace is updated (Fig. 1). Special-purpose processors called 'attention codelets' will monitor the CSM for relevant structures, and bring those structures to the Global Workspace. Each attention codelet has its own particular concern. That is, each scans the CSM for structures that best match its specific, narrow concerns. The best-matching structures are brought from the CSM to the Global Workspace where they compete to become the content of the conscious broadcast.

Coalitions are formed containing attention codelets and various structures. When attention codelets find structures of interest to them in the CSM, they can bring those structures to the coalition-forming process. Here the structures are formed into full coalitions with a combined total activation, and can now compete for consciousness as one object within the Global Workspace. Whatever coalition is the most salient (i.e., has the most activation) wins the competition for consciousness, and its content is broadcast globally across the model to almost every module. Often, it is the brightest, loudest, or most urgent structure that wins the competition for conscious broadcast. With each broadcast, all memory modules can be updated, that is they can *learn*. After most broadcasts, *an action selection process is initiated*. Perception and action are intimately connected because almost every cognitive cycle leads to some action even if that action is minute or internal.

For an overview of the LIDA model see Fig. 1. The cognitive cycle begins to the left with incoming internal and external sensory stimuli, and proceeds roughly clockwise around the model.

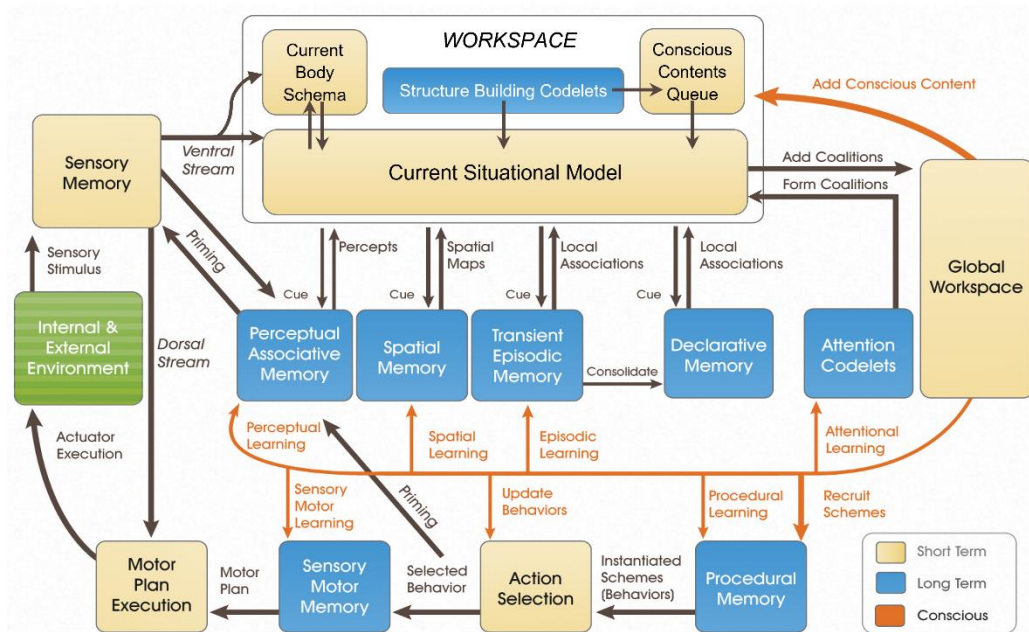


Fig. 1. Overview of the LIDA Model

3. The Narrative Self and Action Selection

Several authors have suggested that what characterizes the self is the agent's ability to create, from its own experiences, a coherent life narrative. Simultaneously, cognitive science, phenomenology, psychology, anthropology, and numerous other fields have suggested and specified various *aspects of the self* (e.g., sense of agency, sense of ownership, preconscious self, minimal self, bodily self, habitual self, and many more). Rather than arguing that the narrative self is *the* true self, we argue that what we call *the self* is the interplay between many dynamically interacting parts, or what Gallagher [2013] has called a 'pattern theory of the self.' The pattern theory of the self has previously been incorporated into the LIDA model, including the autobiographical self and the narrative self [Ryan *et al.*, 2019]. The important point here is that several factors shape the way agents experience themselves as a self, from the constitution of our bodies, the structure of perception, the structure of our memory systems, to the cultural niches we live in. One

product of this list of features is the constant construction of a coherent self-narrative [Dennett, 1992; Schechtman, 1996, 2014]. The self is not a singular thing, but rather emerges from various biological and perceptual processes, which create a coherent self-narrative. From this, we experience ourselves as characters in a story automatically written by us. While many authors have developed versions of the narrative self, the core idea is that temporally the “self” is constituted by a coherent narrative structured from the agent’s memories, beliefs, affect, and bodily capacities [Schechtman, 2014]. Such a narrative not only provides the agent with an existential answer to the question “who am I?,” but more importantly also aids the agent in smoothly understanding and maneuvering in its environment. Hence, a coherent self-narrative is important for knowledge, perception, and action selection.

In elucidating the self, the LIDA model distinguishes between the self-concept and the narrative self. Although having similarities, these two aspects of self serve different functions. Self-concepts include attitudes, beliefs, memories, and emotions related to oneself, among other things. They represent how the agent currently takes itself to be. For example, an agent might believe that it is a charming person, that it is healthy, or that it once went on a TV game show and won. The narrative self, on the other hand, explains why an agent did what they did [Ryan *et al.*, 2019]. In other words, the narrative self is concerned with temporality and causality. Self-narratives provide agents with an understanding of why they acted as they did.

Furthermore, self-narratives aid agents in understanding and acting on their environments and by providing background reasons for intention-setting. For example, when faced with the choice of continuing to pursue a difficult aim, an agent can rely on their self-narrative. “Should I continue pursuing this difficult aim?” “Yes: I am the kind of person who gets stuff done no matter the obstacle;” or no, “I am a downtrodden victim of society’s systemic injustices.” Here both the hero and victim characterizations can be extended into a longer, coherent narrative incorporating the agent’s memories. Unlike simple sequential organization, a narrative organizes an agent’s memory in terms of importance and causality. That is, narratives tell us which life events are important to us and how these events impact later significant events. In this way, narratives also aid in the action-selection process, especially when it comes to deliberation.

Readers and critics of self-narrative theories often misunderstand the notion, because they assume that we are always consciously aware of, and volitionally adding to, our narratives. However, the notion of a self-narrative should not be understood in such a strong sense. Rather, Schechtman [2014, p. 100] argues that self-narratives are mostly implicit, and in LIDA terms, preconscious. For Schechtman, as well as LIDA, the agent’s self-narrative mostly operates at the sub-personal level but can, if needed, be brought to consciousness. However, most of the time self-narratives function as a holistic structure that shapes how we experience, categorize, and act on future events. For example, in LIDA, an agent’s feelings and desires can augment its options differently depending on whether the agent takes themselves to be a “the person who gets stuff done” or “the tragic victim of a cruel universe.”

In this way, the LIDA model can explain how narratives can make previously opaque insights explicitly aware to the agent. In LIDA, through a stream of cognitive cycles, a narrative is generated via mostly sub-personal processes. However, being stored in the agent's memory systems parts of the narrative, when promoted the right way, can be called to consciousness. Through mostly sub-personal processing, the LIDA model preserves the insight that agents generate self-narratives while most times being unaware of the self-narrative.

There often is a consistent, self-reinforcing feedback loop between the narrative schemata an agent has constructed for themselves, and their choices and experiences. For example, the agent who has constructed a pessimistic "tragic victim" narrative will focus on negative experience, and generally cast their experiences in a negative light, thereby reinforcing their self-narrative. Similarly, more positive-oriented narratives tend to cast perceptual experience in a more positive light, and to generate choices that reinforce the positive narrative.

Self-narratives not only aid in creating a temporally-coherent experience of oneself, they also assist in structuring action selection. In particular, deliberative action selection is determined not only by our distal intentions, but also by our personal narratives. For example, the question of whether to use one's disposable income this month on concert tickets, or to put it into savings for a new car, is not simply a cost-benefit analysis. The agent may very well think that the concert is as "beneficial" to their life as adding a little extra to the car savings fund. The question to save or splurge becomes a larger question of how these purchases fit into the agent's self-narrative. "Am I the kind of person who values this type of experience more than savings?" Rather than simply engaging in utilitarian cost-benefit analysis to obtain coherent structured action, agents must often consult their self-narrative. In LIDA, we use an updated ideomotor theory [James, 1890/2007] to account for the role of narratives in action selection. As we will see (Sec. 4.4), narratives partake in the ideomotor process by providing the agent with supporters and objectors.

MacIntyre [2013] elucidates how the structure of behavior is dependent on intentions, and how intentions are, in turn, dependent on self-narratives. In line with prominent figures in action theory [Anscombe, 1979; Davidson, 1985; Grice, 1972], MacIntyre points out that any action can be described in any number of ways. For example, when we see a man doing something in the backyard, is he "gardening," "exercising," "pleasing his spouse," or some other activity? [MacIntyre, 2013, p. 206]. To figure out the man's intentions, it is not enough to have a physical description of what he is doing; rather, we must know the context within which the man is performing the action. In other words, to understand agential action we need to understand the narrative within which the action is embedded. For example, a young boy robot-dancing at a bus stop seems unintelligible without any background. However, if we know that the boy has made it his life's mission to become internet famous because his teacher told him he would never amount to anything, then the spectacle becomes intelligible. As MacIntyre puts it,

We cannot, that is to say, characterize behavior independently of intentions, and we cannot characterize intentions independently of the settings which make those intentions intelligible both to agents themselves and to others [MacIntyre, 2013, p. 206]

MacIntyre's point is that *intelligible action* exists in frameworks where short-term intentions cohere with long-term intentions in a network that adds up to a narrative. Why is the man gardening? Because he intends to surprise his spouse. He intends to do so because he wants to continue their happy marriage, and the happy marriage is a happy one exactly because of key actions and intentions that came before it. Why is the boy pretending to be a robot at the bus stop? In the hope that someone will film his behavior and put it on the internet. Intelligible actions exist in such a way that they form a temporal and causal story. MacIntyre provides a fun and apt example:

If in the middle of my lecture on Kant's ethics I suddenly broke six eggs into a bowl and added flour and sugar, proceeding all the while with my Kantian exegesis, I have not, simply in virtue of the fact that I was following a sequence prescribed by Fanny Farmer, performed an intelligible action [MacIntyre, 2013, p. 209].

We see then that actions are opaque to others and ourselves unless we understand intentions as embedded into narratives. To understand the actions of an agent, we need to understand which of the many possible intentions that led to their action is the primary intention. But primary intentions are only those that make sense when coherently embedded into a narrative. For the man gardening because he wants to continue the happy marriage, the short-term intentions to pick up the shovel and to dig in a certain location are only intelligible considering the history of the marriage. Cracking six eggs into a bowl during a Kant lecture is not an intelligible action unless we can create a coherent narrative of intentions, events, and contexts that led up to the act. In this way, we see that part of what it means to be an autonomous agent is to be the *author* of one's actions. To make sense of our own actions and the actions of others, we are expected to be *legible*.

At this point we have seen how there is a deeply coherent structure in place between self-narratives and our actions. Furthermore, we have seen that intentions are important for understanding, individuating, and causing agential action. A multi-tiered structure is now in place in which self-narrative aids in setting and adjusting distal intentions. In turn, distal intentions structure the setting of short-term intentions and the execution of actions [Bratman, 1987, 2007]. However, as agents go through the world acting on their environments, new experiences, memories, and skills also update their intentions, and self-narratives. Narratives, intentions, and experience exist in both top-down and bottom-up causal relationships.

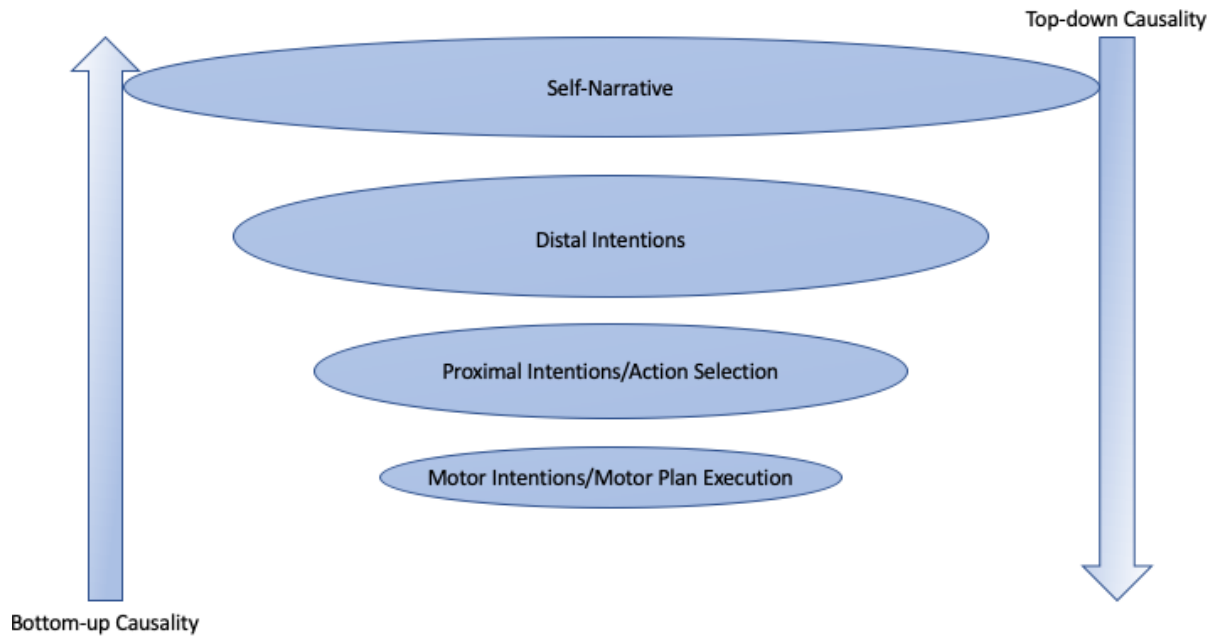


Fig. 2. The coherence relationship between levels involved in agential action. An agent's self narrative aids in creating its distal intentions (its long term goals). In turn, distal intentions help steer more immediate action plans and action execution. When all goes well, there is a coherent, transitive relationship between an agent's most immediate actions and their self-understanding. Furthermore, as agents interact with the world, their self-narrative and distal intentions might change accordingly. In this way, agents and their environments perpetually transform one another.

In action theory, the relationship between high-level, narrative-based intending and our moment-to-moment actions is captured by Pacherie's DPM model – distal intentions, proximal intentions, and motor intentions [Mylopoulos & Pacherie, 2018b; Pacherie, 2006, 2008]. Distal intentions (D-intentions) are the far-reaching intentions that organize our lives over long time scales (e.g., I want to get a Ph.D., or I am getting the house ready for next winter). Proximal intentions (P-intentions) are mid-range intentions (e.g., I intend to make cereal, or I intend to pick up the paint can in the garage). In Pacherie's model proximal intentions typically contain concrete instructions for physical action. In contrast, in LIDA those instructions are contained within motor intentions. In the DPM model motor intentions (M-intentions) are short-term intentions during action that guide the execution of action (e.g., I intend to close my hand at this specific angle to catch the ball). These intentions form a causal hierarchy which is at play whenever an agent acts in the world [Mylopoulos & Pacherie, 2018b].

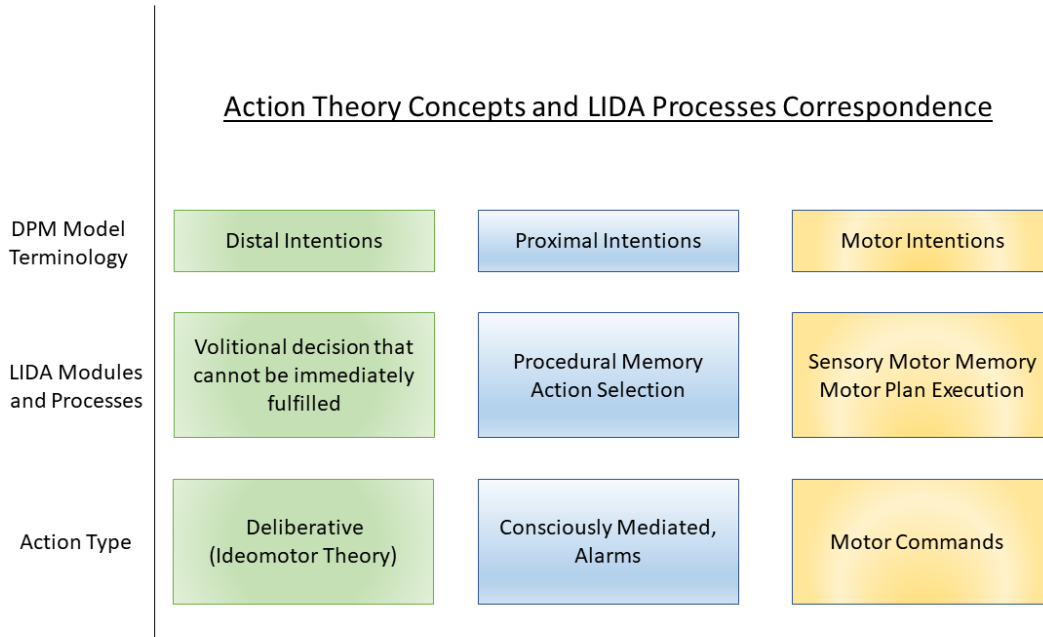


Fig. 3. The figure shows the overlap between terminology from Pacherie’s DPM model and its corresponding modules, and processes in the LIDA model. Note that the LIDA model separates D-Intentions, P-intentions and M-intentions into sub-modules and processes. Thus, the LIDA model operates at a higher level of explanatory granularity than typical action theory.

While the general gist of the DPM model maps well onto the LIDA model, the LIDA model divides some of the intentional processes into finer granularity (see Fig. 3). For example, as we will see M-intentions are divided into two modules and processes (Sensory Motor Memory and Motor Plan Execution) and similarly P-intentions, our everyday immediate actions, are divided into two processes and two modules (Procedural Memory and Action Selection). Hence, the biggest mapping difference between the DPM model and LIDA, regards P-intentions. In the DPM model our everyday P-intentions (making cereal or putting on the green shirt) contain instructions on *how* to do the action. In contrast, as we will see, in LIDA instructions for *how to do an action* is captured in Sensory Motor Memory which chooses motor plan templates and instantiate those into a motor plan (see Sec. 4).

In LIDA, P-intentions are mediated by consciousness. This means that P-intentions come about when Procedural Memory uses content from the conscious broadcast to instantiate behavior schemes and then send them off to Action Selection in a *behavior stream*. In the stream each scheme contains variables for a context, an action, and an expected result. For example one scheme might contain “From the clothes pile on the chair, pick up green shirt, resulting in having a shirt available in hand.” The very next scheme in the stream might contain “In the bedroom holding a shirt, put shirt over your head and pull it on, resulting in having a shirt on.” We see then, that the P-intention to “put on the shirt” contains a series of behaviors each of which must be selected. Thus, P-intention is the

process in which a stream of instantiated schemes (what we call behaviors) is sent from Procedural Memory to Action Selection, which must then pick the most appropriate behavior given the situation (we will explain this process in more detail in Sec. 4.3). Importantly, while the DPM model speaks of “a” proximal intention as a single item, in LIDA P-intentions are *behavior streams across modules*. Thus, the intention to “get the paint from next door” is in reality a series of behaviors that are sent from Procedural Memory to Action Selection. Action Selection then, typically chooses from this series, usually in the sequential order in which they were delivered, which actions are to be further concretized by Sensory Motor Memory, and then carried into physical movement.

However, since P-intentions, are streams of behaviors which are chosen behavior by behavior in Action Selection, the module can stop choosing behaviors from the behavior stream to do something else. For example, if an agent is typing an email, and the doorbell rings, the agent does not have to finish carrying out the behavior stream. Rather, the agent can go answer the door, and then return to their desk to continue the email writing (since the rest of the email writing behaviors from the email behavior stream still linger in Action Selection). Thus, the LIDA model captures the flexibility of everyday proximal engagement with the world that is often interrupted or augmented on the fly, by separating proximal engagement into behavior streams and the action selection process.

In LIDA to *commit* to a proximal intention is a distributed process that typically begins with an option (Sec. 4.2), and involves Procedural Memory, Act

ion Selection, and Sensory Motor Memory. Having committed to a P-intention, means that some consciously broadcast content has been instantiated into a stream of behaviors, and those behaviors are now highly likely to be chosen by Action Selection one by one (unless something more pressing emerges in the environment). For example, when the agent intends to “go get the paint next door” a series of behaviors such as walking, door opening, paint getting might get sent to Action Selection, and are likely being selected and executed in sequential order (again, unless something interferes with the process. For example, the house cat might jump in the way and demand petting). Thus, proximal intentions, rather than being a discrete item, involve a series of processes and modules in LIDA (Sec. 4.3).

In contrast to P-intentions, D-intentions are those intentions that cannot be immediately executed and for which a behavior stream typically cannot be created. For example, creating a behavior stream for getting paint in the garage is fairly straightforward because the agent’s Procedural Memory can reasonably lay out all the behaviors needed to go next door and get the paint. However, no such behavior stream can be created for “getting a Ph.D.” Getting a Ph.D. is a goal that cannot be fulfilled immediately, *and* there is no concrete stream of behaviors that can lead the agent to that goal. Rather, to be fulfilled, the distal intention must be stored in the agent’s memory systems, and be revisited as the agent goes through life. An agent can create P-intentions involved with activities such as writing papers or signing up for classes, but no single behavior stream can typically fulfill a distal intention such as getting a Ph.D. Much of this paper addresses how distal intentions can be revisited and help guide agent’s daily behavior.

As we have seen, action theory tends to conceptualize intentions at a fairly coarse level of granularity. We elaborate on that tradition by taking each kind of intention (D-intention, P-intention, M-intention), to be a heuristic for finer grained processes across various modules. Intentions then are a shorthand for describing cognitive processes that produce structured actions – namely, agential actions.

Many intentional actions in LIDA agents are mediated by the content broadcast from the Global Workspace. However, while consciousness plays a role in distributing content to the various modules so they can fulfill their function (for example, selecting schemes, instantiating behaviors, selecting motor plan templates, etc.) these processes almost always remain unconscious. Thus, the various steps of action selection might not happen in explicit awareness. For example, when deeply emerged into a game of tennis, an agent might end up performing the “volley” rather than the “forehand” but have no explicit awareness of the competition that took place between the “volley” behavior and the “forehand” behavior. In contrast, if an agent is playing Chess, they might engage in an explicitly deliberative process in which much of it comes to phenomenal awareness (Sec. 4.4).

In canonical action theory, action has typically been explained mostly in folk psychological terms -namely in terms of beliefs and desires [Davidson, 1978, 1985]. However, as Bratman points out, beliefs and desires are not enough to form intention and steer coherent behavior. For coherent agential action we also need an evaluation that a certain course of action is desirable and *achievable*: "Intentions are, whereas ordinary desires are not, conduct-controlling pro-attitudes. Ordinary desires, in contrast, are merely potential influencers of action" [Bratman, 1987, p. 15]. For Bratman, having beliefs and desires is not sufficient to throw us into action. Rather, we need intentions as controllers of conduct. Consequently, unlike beliefs and desires alone, intentions involve a *commitment* to action.

As we will see in the following sections, the commitment that Bratman speaks of is implemented throughout various processes across multiple cognitive cycles. Thus, we agree that intentions steer agential behavior through a commitment to the intention. However, commitment is an emergent property that takes place across memory, perception, and learning processes. In particular, the LIDA model is pledged to the conscious learning hypothesis [Franklin *et al.*, 2016, Sec. 4.5]. Thus, learning takes place through the interaction between consciousness and the various memory systems, and the process of forming intentions and commitments is no different. Thus, as we will see, the formation of narratives, distal intentions, and commitments are all mediated by consciousness.

Next, we will show how distal, proximal, and motor intentions map onto LIDA’s cognitive cycle. Furthermore, we will demonstrate how the hierarchy of narrative and the DPM model required for long-term, coherent action is part of the LIDA model. Our implementation also captures Bratman’s insight that intentions include commitments. This, in turn, will demonstrate how LIDA agents can act coherently over varying time scales.

4. The Cognitive Cycle – In A Bit More Detail

The cognitive cycle begins as internal and external stimulation enters Sensory Memory (see Fig. 1). This marks the beginning of the “perception and understanding” phase. In Sensory Memory, low-level feature detectors begin distinguishing the most rudimentary features of the signal. Next, Perceptual Associative Memory (PAM) processes the low-level features from Sensory Memory into usually more complex entities such as objects, categories, relations, events, feelings and emotions. PAM is pivotally important, since a large portion of the content of the Current Situational Model (CSM) is composed of percepts arriving from PAM. As the CSM updates, cognitive maps can be cued from Spatial Memory. Objects and events in the CSM that have been recognized by PAM cue episodes from Transient Episodic Memory and from Declarative Memory (itself containing autobiographical and semantic memory).

As the CSM is updated, special-purpose processors called “structure building codelets” scan the CSM to create simple or complex structures which amongst other things include links between objects, events, and relations. Structure building codelets scan the CSM for material relevant to their concerns, and turn that data into structures within the CSM. Thus, the rich model of the environment that exists in the CSM (complete with affect, memory associations, options, and much more) is perpetually stitched together by structure building codelets. When updating the CSM, LIDA utilizes a nodes and link structure [Franklin *et al.*, 2016]. For example, the event of my friend buttering some toast consists of nodes representing my friend, a butter knife, the butter, and the toast, all connected by relation links to the event node. Similar to attention codelets, structure building codelets only scan the CSM for material that is useful for the structures they want to build.

Finally, the preconscious Workspace also contains a Conscious Contents Queue (CCQ), keeping track of the most recent conscious broadcasts. This helps the structure building codelets create a sense of time and causality [Madl *et al.*, 2015; Snaider *et al.*, 2012]. For example, using the CCQ, structure building codelets can create causal links between events into a temporal causal order.

So far, we have seen how different memory modules create a rich model of the environment and situation within the CSM. However, the Current Situational Model can be far too complex for its entirety to be broadcast to the rest of the system. The agent must *attend* to what is most salient and *act* on that. Hence, we enter the attending phase. To figure out what must be consciously broadcast, the agent utilizes attention codelets. The attention codelet system is a long-term memory system that can learn with each cognitive cycle. Attention codelets each scan the CSM for structures relevant to its concerns, and bring those structures to the Global Workspace. Attention codelets can look for very general and basic attributes of structures such as being loud, bright, large, dangerous, attractive, and so forth. However, given an agent’s specific overarching agenda, goals, and expertise, new attention codelets can also be generated to scan for more specific features such as “a good Tennis backswing” or “poorly welded iron-rods.” In fact, agents frequently generate new attention codelets as they learn, act, and become more attuned with their

environment. The competition for consciousness happens not only between physical features of objects, but also between features that are modulated by our experiences, capacities, and intentions.

As previously mentioned, attention codelets collaborate to form coalitions that compete in the Global Workspace for the spotlight of consciousness. The coalition with the highest activation will be the one whose content is broadcast throughout the system. When some coalition arrives at the Global Workspace and wins the competition for consciousness, the structures carried by the coalition become the content of the broadcast. This content is then sent to every module throughout the model.

4.1. *Affective Valence and Incentive Salience*

Cognition is not a neutral calculation; affect, emotion, and incentive play important roles in the lives of LIDA cognitive agents. To capture this fact about cognition in LIDA, representational parameters such as affective valence (which represents “liking” or “disliking”) and incentive salience (which represents “wanting” or “dreading”) can change the activation value of a coalition, increasing or decreasing its likelihood of being broadcast (see McCall *et al.* [2020]). In other words, perception and cognition are not motivationally neutral, the affective and motivational world of the agent impact what comes to consciousness.

Structures in the CSM such as objects and events can be attached to activated feeling nodes. The activation associated with feeling nodes represents the agent’s immediate hedonic “liking or “disliking of the structure in question. The niceness or nastiness associated with the feeling is represented through a valence sign attached to the feeling node. Here the activation itself represents the intensity of the feeling (how much the agent likes or dislikes the thing in question). Valence signs can be either positive or negative, for example:

a “hungry” feeling node represents a drive, and has negative valence sign, while a “satiated” feeling node also represents a drive but with positive valence sign. A “sweet” feeling node represents a perceptual interpretation of some stimulus with positive valence sign, while a “bitter” feeling node is similar, but with negative valence sign. Intuitively, one may be hungry or satiated in a variety of external situations, and one may interpret a variety of different stimuli as sweet or bitter [McCall *et al.*, 2020, p. 45].

The affective valence of a feeling node is determined by the combination of its activation value and its valence sign. Complex structures, for example seeing one’s childhood home, can contain many feeling nodes, each of which adds to or subtracts from the overall affective valence of the structure.

To illustrate the affective process, we can use the example of a juicy hamburger. In LIDA, if the agent in question is embodied, bodily states, such as hunger, thirst, and

temperature are monitored using homeostatic receptors. Feeling nodes in PAM can be activated based upon these bodily states. In this case, the agent might be experiencing hunger, which activates a feeling node with a negative valence sign. However, as a juicy hamburger is put on the table, it gets recognized by PAM and a positive feeling node is attached to it. Due to the agent's current state of hunger, the hamburger has a higher affective valence and therefore also an overall higher incentive salience.

In LIDA incentive salience represents the motivational attraction or repulsion, that is, the degree of wanting or dreading. Objects have incentive salience because they previously have been liked or disliked, or the agent has learned that the object results in positive or negative changes in its homeostatic states (such as hunger and thirst). Because the agent has liked hamburgers in the past, the hamburger can have a fairly high base-level incentive salience. As agents repeatedly come in contact with a given structure, the liking or disliking of that experience slowly updates the structure's base-level incentive salience.

In our example, because the agent is hungry, the "current incentive salience" is also increased. In LIDA current incentive salience is determined by the current state and situation that the agent finds itself in. Thus, being in a state of hunger means an increase in the hamburgers current incentive salience. The total incentive salience of a structure (such as the juicy hamburger) is the base-level and current incentive salience combined. In LIDA, many events in the CSM (for example, a hamburger is put on the table) are assigned negative or positive incentive salience. This value depends on how much the agent wants the event to take place or how much the agent wishes to avoid the event. Many recognized events in the CSM have some level of wanting or dreading expressed through incentive salience. The role of affective valence and incentive salience becomes important later as we look at the role of distal intentions and narratives in structuring long-term action. The total incentive salience of a structure (base-level plus current incentive salience) and the total activation of that structure (base-level and current activation) contribute to its salience, making it more likely to come to consciousness [McCall *et al.*, 2020].

4.2. *Options*

When elucidating the nature of long-term agential action, it is important to address how an agent comes to have options. In the Current Situational Model, structures are labelled either "real" or "virtual." Real content originates from an agent's sensors, whereas virtual content is created internally. For example, dreams, memories, and imaginings are all virtual events. One specific kind of virtual event is of interest here, namely options.

As agents act on and perceive their environments, potentiality for action will be perceived in that environment. In LIDA, options are "choice alternatives" constructed as virtual events in the CSM. Like all other structures in the CSM, options can have activation and incentive salience. Options are generated in the CSM, and brought to the Global Workspace for competition by attention codelets. However, unlike simple perceptions of objects in the environments, options help guide the action selection process more directly.

For example, upon perceiving logs of wood and a fireplace, an option could be generated in the CSM that the agent could make a cozy fire. This virtual event might have high incentive salience since the event node also might have cued positive childhood memories from Declarative Memory. The fact that the virtual feeling-warm event has high affective valence in conjunction with the agent being cold may give the event high incentive salience. This, in turn, modulates the activation value of the coalition(s) containing the option to make a cozy fire. If a coalition containing the option is brought to the Global Workspace, it might win the competition for consciousness. Having broadcast the content of the coalition, the option to make a cozy fire, the “learning and action” phase begins. As we will see next, when an option has been chosen through a deliberative process, the option is transformed into an intention.

4.3. Action Selection in LIDA

For the purposes of this paper, how an agent chooses and produces action becomes critical. The action selection and learning phase in LIDA roughly corresponds to the proximal and motor intentions in the Pacherie’s DPM model. However, as previously mentioned rather than intentions being just discrete items, they are also a shorthand for describing processes. In the LIDA model, instantiating, selecting, and carrying actions into fruition is a multistep process that is typically mediated through a conscious broadcast (with a few exceptions such as alarms). The LIDA model employs four types of action selection: volitional, consciously mediated, automatized, and alarms. In this paper, we focus mainly on volitional action selection and consciously-mediated action selection. For a treatment of alarms see [Franklin *et al.*, 2016; McCall *et al.*, 2020]. Furthermore, automatized action selection is still being developed in LIDA, and hopefully will be addressed in future papers.

Once the content of a coalition has been broadcast, the content of that broadcast reaches Procedural Memory, where it activates relevant schemes. Schemes are long-term stored templates for what to do in various contexts to achieve various results. Each scheme contains a context, an action, and a result. Each scheme also has a base-level activation which quantifies how reliably the scheme’s action leads to its expected result when taken within its context. Schemes can be, and often are, generalized templates that are then filled in and specified using the broadcast of conscious content. The binding of specific values from the conscious broadcast to a scheme’s variables is called instantiation, and schemes whose variables are bound in this way are called “instantiated schemes.” Instantiated schemes are also referred to as “behaviors,” and these are sent to Action Selection. In many situations, Action Selection may have multiple possible behaviors from which to choose. For example, when dealing with an approaching tennis ball, “volley,” “smash,” or “ground stroke” may all be possible actions that could be encoded in behaviors.

Often a behavior stream is sent from Procedural Memory to Action Selection. The P-intention that resulted in the instantiation of that behavior stream will likely be the result of the final behavior in the stream. For example, the proximal intention of “putting on that green shirt from the pile on the chair” can be used to instantiate a behavior stream

containing behaviors such as picking up the shirt, turning it the right way, putting the shirt over one's head, and sliding into the shirt. Each of these instantiated schemes can be sent to Action Selection in a stream so that they can compete to be chosen by Action Selection one by one. This process ensures that when an agent commits to an intention, they can still change their actions as the action is taking place should something change in the environment. For example, if an agent is in the midst of putting on the green shirt but realizes that there is a big hot sauce stain on the shirt, they do not have to continue putting on the shirt. The hot sauce stain has come to consciousness, and Procedural Memory has now sent Action Selection a new behavior that wins out the competition – namely, put the shirt back down. By having each behavior in a behavior stream compete in Action Selection, the agent can accommodate events that might change the proximal intention they are in the process of executing.

Within Action Selection, the behavior that fits the current situation the best is selected and sent to Sensory Motor Memory. Much like the competition for consciousness in the Global Workspace, behaviors compete in Action Selection, before the winner is sent to Sensory Motor Memory. Sensory Motor Memory is composed of *motor plan templates*, one of which then becomes instantiated using the selected behavior [Dong & Franklin, 2015]. The system creates a motor plan that most optimally fits the selected instantiated behavior. Once the behavior has been paired with a motor plan, the action is carried into fruition in Motor Plan Execution. The execution of the motor plan occurs while inputs from the dorsal stream create rapid, online updates facilitating the execution of the motor plan. For example, if an agent is catching a baseball, rapid dorsal stream updating bypassing the conscious cycle goes directly to motor plan execution to facilitate the smooth unfolding of the motor plan, enabling the ball to be caught effortlessly.

To ensure coherence between our actions, once a behavior has been selected, a special short term attention codelet is generated to scan the Current Situational Model for the expected results or absence of results of that behavior. This attention codelet is called an *expectation codelet*, and it looks for structures in the CSM that are relevant to the expected result of the action to be executed. Expectation codelets (as with all attention codelets) look for structures matching their concerns. When such structures are found, the expectation codelets bring themselves and the structure to the coalition-forming process. These coalitions are then brought to the Global Workspace to compete for consciousness. Expectation codelets and other attention codelets can contribute to the total activation of a coalition making it more likely to win the competition for consciousness. In this way, coalitions that are related to the results of the agent's actions tend to have slightly boosted activation when competing for consciousness in the Global Workspace.

From the expectation codelet-forming process, we see that an agent's behavior can generate some of the conditions of its future patterns of attention. This becomes particularly important when we look at the self-reinforcing role of distal intentions and narratives in relation to attention (Sec. 5.4).

So far, we have seen that action selection and execution in LIDA is a multistep process that moves from the general choice of schemes, to more specific behaviors, to concrete

motor execution; a process that typically involves P-intentions and M-intentions. However, it is important to keep in mind that this process of action selection is typically mediated by consciousness, but is not itself conscious [Franklin & Baars, 2010].

Furthermore, alarm events bypass the conscious broadcast altogether, and move straight from PAM with perhaps a stop at the CSM to Procedural Memory. This bypassing of the conscious broadcast explains why sometimes agents do not realize what they are doing in dangerous situations until after the fact. As an example, we can think about the experience of being cut off in traffic by a hazardous driver; only after hitting the brakes and turning the wheel does the driver typically become fully aware of what has just happened. Thus, in LIDA alarms are never conscious, but still intentional. That is, alarms are intentional because they still utilize Procedural Memory, Action Selection, and Sensory Motor Memory to create actions. In other words, alarms still generate P-intentions and M-intentions. Finally, as previously mentioned automatized action selection is still being developed in LIDA but we suspect that these will also be intentional in the sense that they generate P-intentions and M-intentions.

As we will see next, in the case of volition, the action selection process itself becomes more explicitly conscious. Furthermore, we are now ready to see how distal intentions, and self-narratives help facilitate the action selection process. As we will see, our personal narratives of “who we are,” and our distal intentions, are relevant both for consciously-mediated action selection as well as for volitional action selection.

4.4. *Volitional Action, Ideomotor Theory, and the Self Narrative*

Volitional decision making is often required of cognitive agents. While consciously-mediated action selection is not itself conscious, volitional action selection often is partly conscious. Volitional decisions are those which require the agent to “mull over” or actively deliberate. Volitional decision making is often required in contexts in which the situation might not provide a clear-cut path of execution, the situation is completely novel, or perhaps two or more behaviors could seemingly achieve the desired goal equally well. When such situations occur, the LIDA model employs an updated ideomotor theory [James, 2007; Shin *et al.*, 2010]. However, as Bratman reminds us, in order to break everyday stalemate cases, we must be able to consult, within deliberation, a nested hierarchy of long- and short-term intentions [Bratman, 1987]. Distal intentions and narratives become important pieces in the dance between proposers and objectors of LIDA’s ideomotor process.

Volitional action is characterized by an internal “debate” between proposers, supporters, and objectors. In LIDA, volitional decision making is partly conscious, and importantly takes place over *multiple cycles*. The process of deliberation begins as an option, question, dilemma, or novel, indeterminate situation comes to consciousness, calling for deliberation. As various instantiated schemes compete in Action Selection, the *deliberation scheme* must win the competition by having the most activation in order for deliberation to begin. Typically, this happens when no competing behaviors have high

activation, in routine cases such as looking at a restaurant menu, or when the agent must consider an internally generated proposal. When deliberation is selected, a special purpose processor, called a ‘timer,’ is created and sent to the Current Situational Model. The timer is crucial for volitional decision making because it ensures that there will be an end to the deliberation processes that might otherwise persevere [Franklin, 2000].

As the agent continues to perceive the situation, supporters, objectors, or new proposers can be created in the CSM. If no objection wins the competition for consciousness, and the timer runs out, the choice in question becomes the agent’s immediate goal, and a relevant behavior is selected by Action Selection. This instantiated scheme is then sent to Sensory Motor Memory for the selection of an appropriate motor plan, and the action is carried out by Motor Plan Execution.

However, most deliberative processes include the back-and-forth between supporters and objectors. Thus, if an objector is generated and comes to consciousness, the timer is stopped. However, if a new supporter arises in consciousness, the timer is restarted, this time with less remaining time on it to ensure that eventually a decision will be made. In this way, as each supporter or objector or new proposer comes to consciousness, deliberation takes place over multiple cognitive cycles.

During the deliberative process, special attention codelets are generated. Each of these new attention codelets scan the CSM for structures that are relevant to their concerns. The attention codelets scan for structures that can function as objections or supports for the decision being deliberated. In this way, the concerns of the new attention codelets are directly tied to the decision the agent is mulling over. For example, a novice agent is playing chess and is considering whether or not to move their knight. Having spent some time with an instructor the agent has been taught the maxim “having a knight on the rim is grim.” Learning the maxim will result in the creation of an attention codelet whose concern is knights on the edge of the board. This codelet will bring the agent’s attention to the fact that knights are less effective when placed on the sides of the board. This concern could come to consciousness as an objection as part of the deliberation process.

During deliberation, with the generation of special attention codelets, the agent can simultaneously generate special structure building codelets. Any new structure building codelets in addition to any previously learned structure building codelets can build structures in the CSM relevant for supporters and objectors. Further, these structure building codelets can also build new proposals, that the agent can consider should they come to consciousness. Here the relevant structures that structure building codelets utilize include, but are not limited to, relevant snippets of the agent’s self-narrative and distal intentions. During deliberation, structure building codelets build possible objectors and supporters, and attention codelets try to bring these structures to consciousness.

For example, an agent receives an email from a toy company stating that their son’s favorite action figure is back in stock. An option is generated in the Current Situational Model, wins the competition for consciousness and is broadcast to Procedural Memory among many others. Several behavioral streams are instantiated, one being “click-on-link-and-buy,” another “deliberate.” Here the “deliberate” behavior is a deliberation scheme

instantiated to deliberate with the option to click-and-buy as a proposer. In this example, the deliberation behavior wins the competition in Action Selection. As the agent begins the deliberation process, relevant portions of their narrative and distal intentions are cued into the CSM, where structure building codelets use them to make objectors and supporters, and attention codelets bring coalitions including those to the Global Workspace. In this case, a portion of the agent’s self-narrative, “I want to be a loving parent, because my own parents were stern and cold,” becomes a support structure and wins the competition for consciousness.

However, during the very next broadcast (before the timer runs out), structure building codelets find distal intentions set by the agent to not support companies who use exploitative labor practices. This distal intention is further connected to the parts of the agent’s narrative that cast them as someone who wants to be a better example for their son than was their own father. In this case, the narrative segment to be the kind of person that will not purchase from exploitative corporations may be built into an objector by structure building codelets, and may win the competition for consciousness. If the objector wins its way to consciousness, then the running timer would be stopped, and the option to purchase the toy would no longer be a live option for the agent.

Importantly, in the LIDA model, when the ideomotor process converges on a decision, the option the process decides on is transformed into a goal for the agent. As mentioned at the beginning of the paper, we here use ‘goal’ and ‘intention’ synonymously. Intentions are generated at the end of the ideomotor process. Furthermore, as the option is transformed into an intention, the system also generates an intention codelet (a type of attention codelet looking for structures relevant to the intention). This becomes significant later as we elucidate how distal intentions bias agential action (Sec. 5.3). However, first we must look at how the self-narrative involved in the ideomotor process comes about.

5. Narratives and Distal Intentions in LIDA

5.1. *The Narrative Templates and How They Are Generated*

We have seen that narratives and narrative snippets can function as objectors and supporters in the ideomotor process and that this role may help steer the agent’s behavior. Next, we will see how narratives are generated within the agent, through the ongoing cognitive cycles.

With each cognitive cycle, Transient Episodic Memory (TEM) typically updates itself with content from the conscious broadcast. Undecayed contents in TEM are periodically consolidated into Declarative Memory. TEM functions as a long-term working memory storing information that is relevant on a fairly short time scale. Longer-term relevant information is consolidated from TEM to Declarative Memory, as either an autobiographical or semantic memory. Here, ‘autobiographical’ refers to episodes that are stored as occurrences that specifically happen to the agent. In contrast, semantic memory

is stored as general information without specifics, such as "when" and "where." In many complex agents there is a difference between autobiographical memories and semantic memories. However, until now this distinction has only been implemented in a few LIDA agents.

In more complex LIDA agents, a self-narrative is slowly generated and updated as information is consolidated into Autobiographical Memory. As agents deliberate, they often cue autobiographical memories from Declarative Memory into the Current Situational Model where they can help act as supporters or objectors in the deliberation process. If a particular memory or set of memories gets cued frequently to aid in the deliberation process, those memories can begin to stand out as important parts of the agent's self-identity. Each time the memory is part of a conscious broadcast, either as an objector, supporter, or simply being remembered, that memory gains added base-level activation as it is re-updated into Transient Episodic Memory. In other words, the more times a specific memory is broadcast, the higher its base level activation will be, and the more important it becomes in the agent's self-narrative.

Through this process, a set of memories in Declarative Memory begins to emerge as the agent's self-narrative. These high-salience memories are, in turn, organized in a causal pattern (a causal story) within Declarative Memory. While stories come in a variety of types, the kind of story that constitutes a self-narrative is the kind that can explain the causal connections between events over longer time scales. Causal stories connect important events despite those events being temporally distant [Nelson & Fivush, 2020]. For example, graduating college, getting a Ph.D., and being sent to the International Space Station are typically temporally distant events that can be organized in a causal order without having to list all the actions and events in between them. As specific memories gain an increase in base-level activation, those memories can get chained together in a self-narrative that forms a causal story within Declarative Memory.

However, merely gaining an increase in base-level activation does not explain how memories get causally chained together. The chaining together of salient memories takes place across multiple cognitive cycles, and is dependent on the Conscious Contents Queue and *narrative templates*. Human agents are frequently exposed to narratives. Through fiction, folklore, TV, movies, songs, dance, plays, video games, and through many other cultural and interpersonal means, humans are presented with narratives on how to act in various situations and across slow and fast time scales. These socially- and culturally-transmitted narratives can function as templates for the agent to structure their own life story, their self-narrative [Bruner, 1991; 2004]. As Bruner aptly reminds us, quoting Oscar Wilde, life often imitates art [Bruner, 2004, p. 692]. Furthermore, from early infancy, human agents frequently interact with their caretakers through narrative interaction [Delafield-Butt & Trevarthen, 2015]^c. Culturally and socially transmitted narratives become templates that aid agents in organizing their experiences into self-narratives

^c Narratives are so pervasive in human culture and social development that some authors argue narrative competency plays a critical role in the development of theory of mind [Gallagher & Hutto, 2008; Hutto, 2008].

[Nelson & Fivush, 2020]. Without such templates, making sense of oneself and the world on longer time scale, involving extended periods of time over months and years, would become increasingly difficult for the agent.

Whether through interaction with culture, media, or other agents, human agents are presented with a myriad of narratives. Over time, such narratives are consolidated from TEM to Declarative Memory. As the agent continues to engage with the world, these narratives often begin to lose their specificity (who, where, when, and why?) over the course of many cognitive cycles. This loss of specificity typically happens as some of the details decay away over long periods of time. In some instances, the details are no longer relevant and simply decay away from the semantic information. In other instances, new memories containing the same semantic information interfere with time and place details of the original memory. For example, if an agent is told by Fred that Nashville is the state capital of Tennessee, they might over time forget that Fred had told them so, while still remember the fact about Nashville. As the information about Nashville is presented in other situations the semantic knowledge is reinforced while the original details (who, where, and when) are interfered with. The agent can no longer remember where they learned the semantic knowledge (which originally came from Fred); they simply remember the reinforced semantic knowledge (Nashville is the state capital of TN). In a similar fashion, agents often learn specific detailed narratives from their cultures. However, as the details

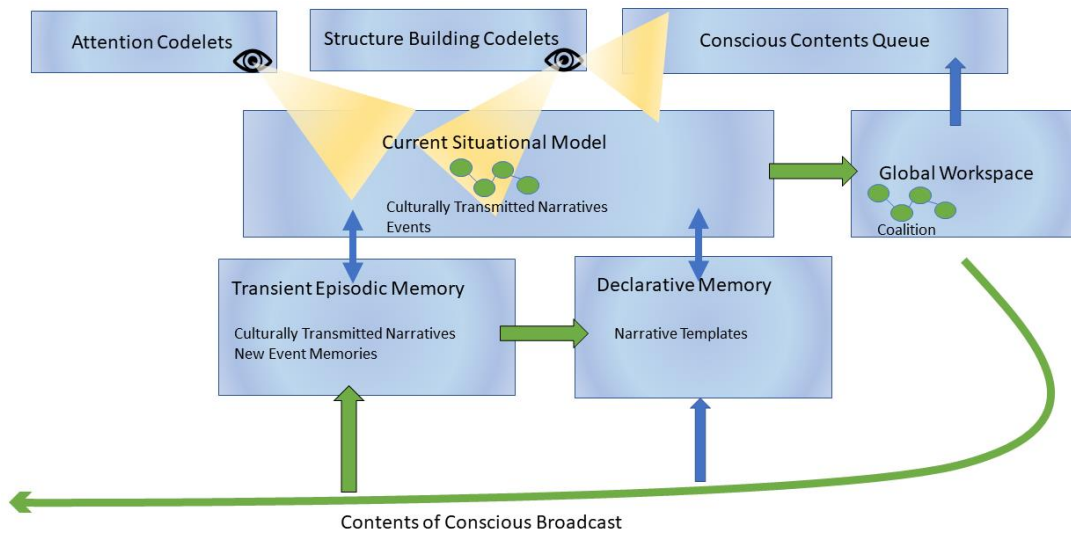


Fig. 4. Culturally transmitted narratives can become part of coalitions that might win the competition to be broadcast. The green figures represent the typical trajectory a narrative might take through the model. Blue arrows represent potential cuing of data structures or updating. Structure Building Codelets scan the CSM and Conscious Content Queue to build structures, while attention codelets bring structures to the coalition forming process and the Global Workspace. Narratives are often broadcast, stored in Transient Episodic Memory, and later consolidated into Declarative Memory. Over many cognitive cycles narratives can become reinforced and generalized thereby taking the form of a narrative template.

decay away, what typically remains is a culturally-transmitted, generalized narrative template [Freeman, 2007].

In LIDA, the process of turning culturally-transmitted narratives into narrative templates happens through multiple cognitive cycles (see Fig. 4). Memories that have lost some or all their specificity are frequently cued into the Current Situational Model (for example, if the agent encounters something in the environment that reminds them of those narrative memories). If more than one culturally-transmitted narrative memory is cued into the CSM at the same time, structure building codelets can begin to form those memories into a complex structure. A subset of structure building codelets look specifically for similarity in narrative structure for the purpose of narrative template creation. We must remember that codelets are special-purpose processors that scan the CSM only for things related to their one concern, in this case the creation of narrative templates. As structure building codelets find sufficiently similar cued up narratives in the CSM, they begin to form these into narrative templates. Specifically, the structure building codelets are *generalizing* over the narratives, to create a general template in which the specific details of the narrative are now substitutable place holders (variables). For example, with enough narratives such as “Mario saves Princess Peach”, “Spiderman saves Mary-Jane”, “Solid Snake saves Meryl” an agent can generalize the (gendered and often harmful) template male-agent-saves-helpless-female. Through contact with enough similar narratives, structure building codelets can create a variety of generalized templates with unbound placeholders.

If a narrative template is created by structure building codelets, it can become part of a coalition that potentially wins the competition for the conscious broadcast. If the template is part of the content being broadcast, it typically enters TEM and is typically consolidated into Declarative Memory. However, if the broadcasted narrative template is similar enough to a template that already exists in Declarative Memory, then Declarative Memory might simply update the base-level activation of its preexisting narrative template. Learning has occurred.

Furthermore, already existing narrative templates can be adjusted, fleshed out, or changed in other ways as the agent comes in contact with new, similar, culturally-transmitted narratives. If a narrative template is cued into the CSM together with a new narrative memory, structure building codelets can once again do their work and compare the new narrative to existing templates. If the new narrative fits the template, the structure building codelet often will attach it to the narrative template. If this augmented structure is broadcast, it can make it back into TEM. In this way, over many cognitive cycles agents can gain and augment narrative templates. Learning has occurred.

Over time, as agents are exposed to enough narratives, specific templates begin to emerge as they are repeatedly reinforced through the cognitive cycles. In human agents, culturally transmitted narrative templates have been shown to reproduce aspects of self-concept and self-narratives related to race [Collins, 2002], gender [Gryzman *et al.*, 2016; McLean *et al.*, 2020], sexuality [Compton, 2020], and national identity [Lae, 2019]. Furthermore, it has been argued that humans use narrative templates to generate the

narrative self through *self-directed subvocalization* of culturally transmitted narratives. In other words, through *internal speech* as a form of self-stimulation, humans retell narratives originally about others now about themselves [Dennett, 2017]. In this way, culturally transmitted narratives can slowly be transformed into templates that the agent exploits to organize and understand its own experiences. For example, through movies, books, and TV, the narrative of the reluctant hero can become a narrative template for agents to utilize when organizing their experiences. Broad cultural phenomena such as Harry Potter, Star Wars, and Spider-Man all reinforce the reluctant hero narrative. Over time, some human agents might use the basic structure of such stories to evaluate and organize their own experiences. For example, through careful analysis of decades of life stories, McAdams [2006] identifies culturally transmitted templates (or master narratives) that are pervasive in the United States – the underdog, rags to riches, heroism through redemption. Agents organize their experiences to fit the mold provided to them through cultural immersion.

Often, the stories from which agents acquire their narrative templates are far removed from the situations to which the templates are applied. For example, a high-school student gets nominated to be the student representative on the school board. While the student has no desire to be on the board and represent the student body at awkward meetings with teachers, the student might understand that they were chosen to do so. Due to their previous display of excellent public speaking skills, the student is (supposedly) the only one who can do this, the “chosen one.” Even when the original story from which the template was generated is far removed from the situation at hand, agents may nonetheless organize their experiences to fit them.

5.2. *Chaining Together Salient Memories and Templates into Self-Narratives*

At this point, we have two critical pieces of the puzzle: through repeated broadcasts, important life memories gain an increase in base-level activation, and through repeated exposure and memory consolidation, the agent stores narrative templates in Declarative Memory. Next, we will see how memories and narrative templates are chained together to form the agent's self-narrative (for an overview, see Fig. 5).

In many different situations (whether in deliberation or some other cognitive task), salient memories might be cued into the Current Situational Model. Furthermore, often when an agent perceives a situation, that situation might cue up a narrative template from Declarative Memory. For example, an agent might see their friend struggle with and make a hard decision in a moral dilemma (should they or should they not report their colleague to the dean for questionable use of university funds). As the agent talks the case over with their friend, the perceived situation cues up a narrative template that has similar components to the situation of the agent's friend. If the narrative template becomes part of a coalition that wins the competition for the conscious broadcast, the narrative template is temporarily stored in the Conscious Contents Queue. Additionally, as the agent perceives

their friend lamenting over the problematic moral dilemma, salient memories from their past become part of coalitions that win the competition and are broadcast.

In this example, salient life memories and narrative templates are now present in the Conscious Contents Queue (and possibly in the CSM). This allows structure building codelets to stitch the template and the memories together into one combined structure in the CSM. If this structure (templates plus memories) becomes part of a coalition that wins the competition for broadcasting, the narratively-organized structure is broadcast throughout the system. Consequently, the contents of this broadcast, the narratively-organized structure, can be stored in TEM. Later, as TEM consolidates its content into Declarative Memory, the narratively-organized memories can be stored within Declarative Memory. Through this repeated process, across many cycles (sometimes this can take decades in humans), narrative segments become chained together with other narrative segments slowly forming a full causal story. To summarize, as memories and narrative templates in the Conscious Contents Queue and the CSM are chained together by structure building codelets, they gain the chance of becoming part of coalitions that can be broadcast (see Fig. 5 and Fig. 6). If they are broadcast, the narratively-organized memories can be stored and consolidated back into Declarative Memory, where a coherent story is slowly forming.

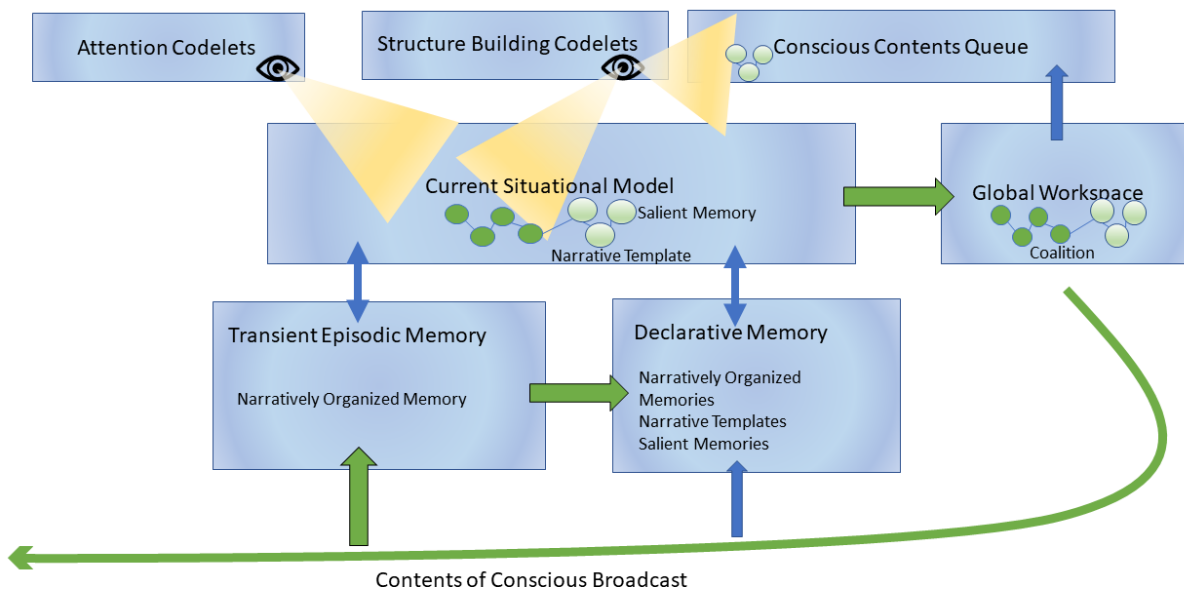


Fig. 5. As we see here the trajectory taken by narratively organized memories is very similar to the trajectory taken by narrative templates. The green figures represent the typical trajectory of stitched-together memory chains. Blue arrows represent possible cuing of data structures or updating. Chains of already narratively-organized memories can be stitched together with salient memories or other chains of memory by structure building codelets. Memory chains already in the CSM or in the Conscious Contents Queue can become part of coalitions and might win the competition for broadcasting. If broadcast, the new coalition can be stored into TEM and later consolidated to Declarative Memory.

The narrativizing process just described can, over time, be repeated, chaining together already narratively-organized memories into larger narratives. A similar process to that responsible for organizing salient memories into narrative form is typically also responsible for taking each of those narratives, chaining them together into a larger narrative. By having templates and narratively-organized memories called into the CSM, structure building codelets can stitch together even larger narratives that can be broadcast and stored in the agent's memory systems. Painting with broad strokes, we see that a self-narrative can be generated through repeated stitching, broadcasting, and memory storing over many cognitive cycles (see Fig. 5 and Fig. 6). In this way we see that the process of creating narrative templates and the process of chaining together self-narratives are in many respects very similar.

It is pertinent to point out that as the self-narrative grows through repeated stitching, chaining, and organizing, it eventually becomes too large for the entire narrative to be broadcast. Thus, it is highly unlikely that agents over a certain complexity and age can ever be consciously "aware" of their entire narrative. Rather, it is more likely that the system generates a form of "outline" or "synopsis" of its own self-narrative that can then be utilized in various processes. Such outlines can represent the whole narrative or sections of it. These outlines are represented in the CSM in the form of a narrative node structure. Similarly, in many cases some of the narrative templates become too large for the system to broadcast the template in its entirety. In those cases, the template becomes represented in the system as a synopsis narrative template node structure. In the same fashion such node structures comprise a summary representation of the full template but are not themselves the full template. In this way neither the agents overarching self-narrative nor its narrative templates put too heavy of a computational burden on the agent, since the agent does not have to hold the entirety of large templates or its entire self-narrative in the CSM.

While episodic memories have to go through TEM before they can get consolidated into Declarative Memory, already-existing narrative memories inside Declarative Memory can have their base-level activation directly updated using contents from the conscious broadcast. Say a portion of an agent's self-narrative becomes part of a coalition that wins the competition for consciousness. The contents of that coalition are broadcast to the agent's various modules. Declarative Memory can use the broadcasted portion of the self-narrative to update the base-level activation of that corresponding portion inside Declarative Memory. In this way, even if elements of an agent's self-narrative are not rearranged or changed, the self-narrative can still be modified through the updating of base-level activation (Remember that base-level activation of that portion of the self-narrative as a structure in the CSM is one component of the coalition activation, which determines which coalition wins the competition for consciousness).

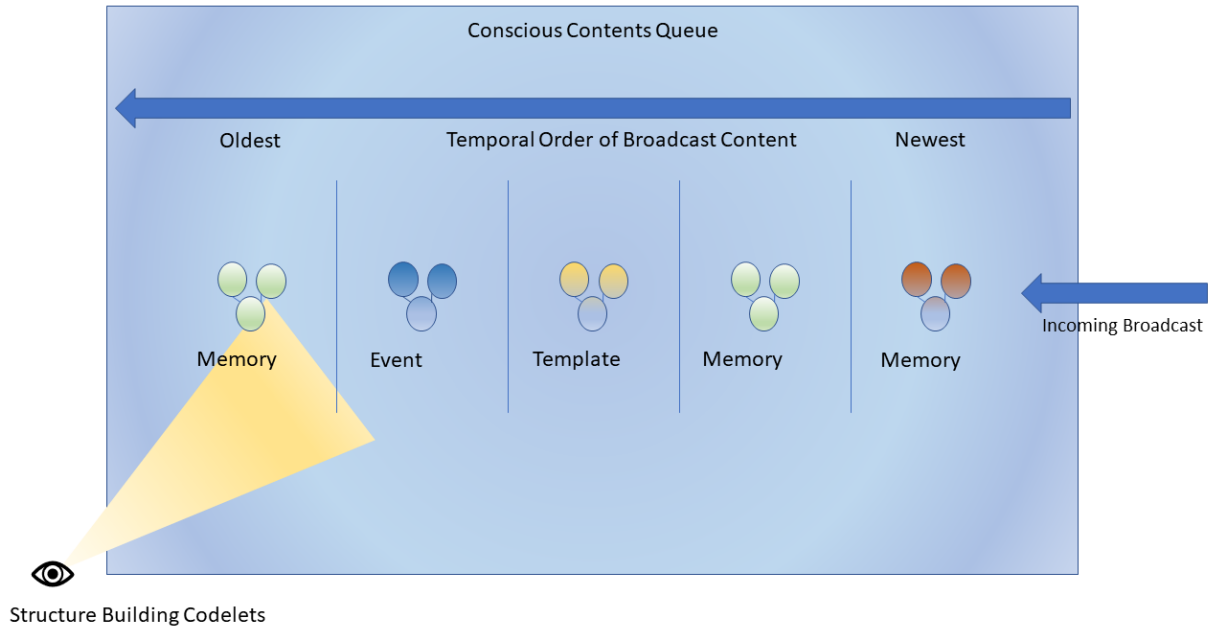


Fig. 6. Here we see an example of content temporally aligned within the Conscious Contents Queue. Since memories and templates are both in the queue, they are available for structure building codelets to bring them into the CSM to become part of structures. In this case, the memories will be organized into narrative form using the available template

While simple agents might have a preprogrammed agenda, more complex agents such as humans have a more "existential" sense of who they are. That is, besides their basic biologically determined agenda (e.g., nutrition, survival, procreation), complex agents develop specific agendas containing, amongst other things, life goals. Wanting to be a great painter or living life my way can, of course, in some reductive fashion be traced back to an agent's simple biological agenda. However, more specific life goals emerge in complex agents from the development of a self-narrative. In LIDA, the agent's agenda is guided by the perpetual subtle reorganizing of the self-narrative in Declarative Memory. However, we wish to stress again that in concert with the pattern theory of the self, the self-narrative is far from the only aspect of what it means for the agent to have a self [Ryan *et al.*, 2019]. Nevertheless, organizing salient memories into a causal narrative provides the agent with a temporal sense of who it is in the world, and what the agent ought to do.

As deliberation is initiated, portions of the self-narrative can be cued up from Declarative Memory into the CSM, where structure building codelets can use them to build objectors, supporters, or proposers. However, not every decision in need of deliberation needs to invoke the self-narrative. Many everyday choices simply utilize concerns embedded in the current situation in order to end the deliberation process. A human agent

may deliberate briefly about which of two varieties of jam to spread on the toast. However, self-narratives still structure the daily, environmentally embedded actions of the agent through distal intentions and the updating of incentive salience. We must remember that many of an agent's distal intentions often stem from the agent's self-narrative and self-concept. Next, we will look at the role of distal intentions and how distal intentions structure action.

5.3. *Setting a Distal Intention in LIDA*

As mentioned at the beginning, in the LIDA literature we have in the past talked about options and goals (Sec. 1 and 4.2). However, here we use the term 'intention' to stay in communication with the literature from cognitive science and action theory. We use the terms 'intention' and 'goal' synonymously. We have seen how narratives are generated and how they play a causal role in long term agential behavior. Next, we will look at distal intentions in more detail to elucidate their role in the hierarchy of long-term agential action. In LIDA there are multiple ways an agent can set and commit to distal intentions. In this paper we focus on setting distal intentions both with and without deliberation.

First let us look at how an agent can set a distal intention through deliberation. Often, an agent will generate an option in response to its environment. (Recall that an option is a virtual event representing a possible choice.) Frequently, if the option comes to consciousness, the ideomotor processes is initiated, making the agent deliberate between whether to act on the option, on some other option, or perhaps not at all (Sec. 4.4). If the option is in fact selected at the end of the ideomotor process, that option is transformed into a goal for the agent – an intention. Thus, if an option comes to consciousness, and a resulting behavior is selected through the Action Selection process and executed, that option, as its result, will be transformed into a goal [Franklin *et al.*, 2016, Sec. 6.2]. However, not all goals can be immediately fulfilled; some goals are in the distant future (e.g., getting a Ph.D., proposing to Sally next year, collecting and assembling all of the 1990 LEGO Castles collection). Distal intentions are goals for the agent that cannot be immediately fulfilled. However, since the goal cannot be immediately fulfilled, the distal intention (through the conscious broadcast) usually ends up being stored in the agent's memory systems. For example, it could be stored in TEM and then potentially consolidated into Declarative Memory if it does not decay away first. Once stored in the agent's memory system the distal intention can affect the agent's behavior through a number of mechanisms (that we will cover in subsequent sections).

Since the distal intention generated at the end of the ideomotor theory is not immediately actionable, it must be "committed to" as something that will be done later. Thus, if an agent decides on an option at the end of the ideomotor process, and that option is part of the conscious broadcast, then based on the option Procedural Memory can instantiate a "commitment" scheme into a commitment behavior. Both the commitment scheme and its commitment behavior contain an *internal* action. Unlike external actions that modify the agent's external environment, internal actions can initiate internal

processes that can operate within the CSM. If the commitment behavior wins the competition amongst behaviors in Action Selection, the result of its internal action in the CSM is a representation of the committed-to distal intention. If that distal intention becomes part of a coalition that wins the competition for consciousness the distal intention is broadcast throughout the system. This means that the agent's memory systems can now store the committed-to distal intention. Thus, deliberating, and then committing to a distal intention is a multicyclic process.

However, many distal intentions do not require the ideomotor process. In other words, agents often set distal intentions without the need to mull over the option. For example, if an agent opens the refrigerator and sees that it is near empty, the option to go to the store tomorrow morning, might come to consciousness. In that case the agent does not need to carefully deliberate over whether or not to go to the store. That intention can be generated as an option in the CSM by structure building codelets. As we shall see, committing to a distal intention without the ideomotor process in LIDA is very similar to committing to one with the ideomotor process. Namely, in LIDA committing to a distal intention without the ideomotor process happens through the use of the same "commitment" internal action.

As in the previous case, if a distal intention comes to consciousness (for example go to the store tomorrow), Procedural Memory can use that distal intention to instantiate a commitment scheme and send the resulting commitment behavior to Action Selection. If that commitment behavior wins the competition in Action Selection, the agent performs an internal action. This action is a commitment action, as in the deliberative case. The result of the commitment action is a representation of the committed-to distal intention in the CSM. If this commitment representation becomes part of a coalition that wins the competition for consciousness, then the content of the coalition will be broadcast, including the committed-to intention. The commitment to the distal intention is thus broadcast and sent throughout the model, where it can be picked up by the agent's memory modules. Once the distal intention has been stored in the agent's memory modules it can begin having an impact on the agent's behavior.

We see then that it typically takes several cognitive cycles for an agent to commit to a distal intention even without deliberation. First an option must become part of a commitment behavior and then the resulting committed-to representation of the distal intention must be part of a coalition that wins the competition for consciousness. Setting distal intentions is always a multicyclic endeavor. Importantly, setting a distal intention involves the agent committing itself to the distal intention so that the intention can guide their behavior going forward. Next, we will look at various processes that help guide behavior once a distal intention has been set

5.4. *Distal Intentions and Attentional Processes*

Once a distal intention has been created, typically at the end of the ideomotor process or through an internal commitment behavior, new structure building codelets and attention codelets are often generated. The concern of these codelets is, respectively, to build

structures in the CSM relevant to the distal intention, and to bring such structures to the Global Workspace to compete for consciousness. Specifically, intention codelets are a subset of attention codelets that look for structures in the CSM relevant to an agent's intentions [Franklin *et al.*, 2016, Sec. 5.4]. When an agent intends to do something over a distal time scale, the process typically involves biasing its own *attentional processes* towards the aim of the distal intention. Consequently, it is an important feature of intention codelets that they decay away at a slow rate. Thus, intention codelets can stay within the system over long time scales, continually scanning the CSM for structures related to the distal intention.

All attention codelets have a base-level activation that can be reinforced during a cognitive cycle. Attention codelets are reinforced in accordance with how successful they are at bringing structures to consciousness. If an attention codelet's base-level activation falls under the minimum threshold, the codelet is deleted. Intention codelets can stay in the system if they are successful at bringing structures to consciousness and are thereby reinforced.

The lingering of intention codelets explains how an agent can go about its day, see something, and, seemingly out of the blue, be reminded of one of their distal intentions [Baars *et al.*, 2007]. Intention codelets related to distal intentions typically are reinforced and don't decay away quickly. When an agent encounters objects and situations related to a distal intention, options related to that distal intention are more likely to come to consciousness. This is so even if the agent has not thought about, or done anything related to, the distal intention for a while. Because of the existence of a distal intention, for example, the desire to become an excellent journalist, an agent's attentional processes are biased towards things related to the achievement of that distal intention. This process may include the generation of new attention codelets, such as an intention codelet (see Fig. 7). For example, by meeting other inspirational people, a preconscious thought about the possibility of becoming an excellent journalist might be generated in an agent's preconscious Workspace. If this possibility wins the competition for consciousness and is broadcast, the possibility can be stored in the agent's memory systems, and may eventually begin to generate codelets. These newly-generated attention codelets may now lead the agent to see actions, events, and options related to the distal intention of becoming a journalist.

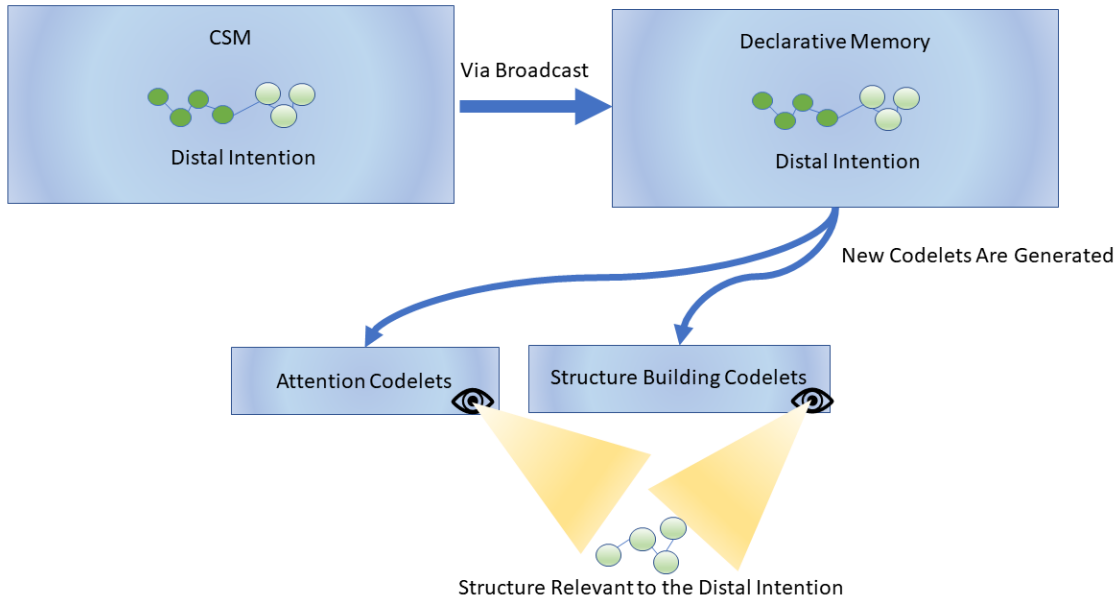


Fig. 7. As distal intentions are created through the ideomotor process - their existence will often prompt the system to generate new attention codelets—specifically, an intention codelet for each new distal intention. These codelets will bring distal intention-relevant structures to compete for consciousness. Intention codelets related to distal intentions decay away at a slow rate, ensuring that the agent can bring structures to consciousness related to their distal intentions even over slow time scales.

However, it is crucial to understand that many processes in LIDA do not have a single cause. Similarly, while distal intentions often aid in the generation of new codelets relevant to their concerns, many other processes can also generate new attention and structure building codelets. For example, actively learning of a new skill, a new academic literature, a new language, board games, or even how to interact with specific agents such as friends and family, can all generate new codelets. Often, the generation of new attention and structure building codelets happens to come from situations that are already in accordance with the aims of distal intentions [Baars *et al.*, 2007]. Thus, while having a distal intention in itself can generate some new codelets, many other processes related to the distal intention also bring about new codelets. For example, pursuing journalism as a major in college means taking classes on writing and investigation, which involves skills that lead to the generation of both new sub-intentions and new codelets. Finally, it is worth noting that while there is a technical story to be told of how new codelets are generated in LIDA, (using amongst other things the Slipnet from the Copycat Architecture [Hofstadter & Mitchell, 1995], that story is outside the scope of this paper.

While in many cases an agent already has all the attention and structure building codelets needed, novel possibilities that the agent has never considered typically require the generation of new codelets. For example, when setting the distal intention to learn how to paint human faces, the novice must generate new attention codelets and structure building codelets that are relevant to the proportions of human facial anatomy. For example, through practice and learning, the novice painter generates attention codelets that scan for subtleties in the ratios of human facial features. Looking for such subtleties was not something the agent engaged in before the generation of the appropriate attention codelets. Setting or gaining a distal intention often involves generating new attention codelets for the fulfillment of the distal intention.

Another common example is learning how to play games. When an agent learns how to play a game such as tic-tac-toe or chess, it is introduced to new pieces, rules, movements, and board positions. Being able to play chess entails understanding the pieces and the significance of their positioning. Learning to play chess is not only to internalize the rules of the game, but also to generate new structure building codelets that are relevant to the game of chess. Developing the distal intention to one day be better at chess than World Champion Magnus Carlsen includes the generation of structure building codelets relevant to chess, and attention codelets relevant to advantageous and disadvantageous board states.

However, gaining a distal intention only once is typically not strong enough to play the conduct-controlling role described above. As the agent goes through events in the world, it might recreate the distal intention as a live possibility in the CSM. Each time the distal intention presents itself and is broadcast, the distal intention in Declarative Memory can be strengthened through an increase in base-level activation. For example, in the case of the agent who realized it could be a journalist, perceiving that possibility once is not enough. Rather, the agent might see journalists in TV shows, fiction, movies, educational advertisement, parental support, and peer suggestions. With each broadcast of individuals who are journalists, the distal intention may gain more base-level activation to the point where it might begin to outweigh other considerations when the agent is deliberating its choices.

However, gaining an increase in base-level activation is typically not enough for a distal intention to guide an agent's conduct. As an agent learns more about what is entailed by their distal intention, new attention codelets are created. In our journalist example, the agent might only have a vague idea of what is actually entailed in being a journalist. However, as the agent learns more about the job, new attention codelets are generated that scan the CSM for things relevant to the new expanded concept of journalist and the expanded distal intention of becoming a journalist. As agents pursue their distal intentions, they typically learn more about those intentions, which in turn provide the agent with more relevant things to pay attention to with regards to the distal intention.

In this manner, distal intentions are often slowly self-reinforcing. As distal intentions are fleshed out, more relevant attention codelets can be generated. When those attention codelets bring coalitions to consciousness that are related to the distal intention, the base-level activation of the distal intention may be updated, thereby making it more likely that

the distal intention will win the competition for consciousness in the future. In this way, when successful in guiding behavior, distal intentions slowly reinforce themselves through a feedback loop of attention and action.

5.5. *Distal Intentions and Incentive Salience Affect Attention and Action Selection*

As shown in the overview (Sec. 3) and illustrated in Fig. 2, there is a hierarchical nature to agential action. However, agential action is not always a top-down process starting with the narrative. While the ideomotor process of deliberation often directly consults the self-narrative, usually the causal power of the narrative is expressed implicitly through distal intentions. Those distal intentions, in turn, aid in the updating of incentive salience.

When a distal intention is stored, each broadcast of a structure that relates to the obtaining of the aim of that distal intention typically results in a positive increase in the structure's base-level incentive salience. Through a history of interaction with events that are related to the aim of the distal intention, the agent builds up a wanting for events that bring it closer to the aim of the distal intention.

For example, an agent might enjoy sweet things and have repeated experiences of eating cakes. Cakes can then become associated with the liking of sweet things and the desire for cake can slowly emerge. In LIDA, the liking is quantified through affective valence and the desiring through incentive salience.

Events that lead to the obtaining of the distal intention can develop incentive salience. If a distal intention has positive incentive salience, and the agent engages with events that are causally related to the attainment of that distal intention, then that event usually also has positive incentive salience. For example, to get a PhD, students need to take a comprehensive exam. That exam may not be inherently likeable, but it may gain incentive salience due to its association with the larger goal of getting a PhD. Similarly, through interaction with events that are related to the distal intention, but are *obstacles* for the agent, the system might decrease the incentive salience of such structures. Therefore, the base-level incentive salience of relevant structures can be increased or decreased with the broadcast of events that relate to the fulfilment of the distal intention.

We see then, that making a self-commitment often means that relevant actions become more salient to the agent. LIDA agents attune themselves to be more incentivized by options, events, and actions that are related to their long-term goals. While commitments are often not conscious, the conscious broadcast plays a role in shaping the degree of our commitments. Being committed to the aim of a distal intention often results in having structures related to the intention slowly gain more incentive salience over multiple cognitive cycles.^d

For example, let us say our aspiring journalist works for their college newspaper. The agent writes an article, and receives positive feedback from the editor. The agent likes

^d For more details regarding motivational processes in LIDA see [McCall *et al.*, 2020]

gaining positive feedback, which increases the incentive salience associated with article writing. Over time, as the agent receives praise from colleagues and friends, the incentive salience of structures related to article writing is increased. Put simply, over time, liking translates into wanting, and wanting something will increase its likelihood of coming to consciousness. As article writing activities gain an increase in salience, this increases the likelihood that the agent will choose these activities over other activities when possible. We see then that there is a relationship between distal intentions and the choice of an agent's day-to-day activities.

6. Summary and Concluding Remarks.

We have looked at the overarching nested structure of narratives and intentions that enable coherent agential action. We have demonstrated how narratives and distal intentions can be generated through cognitive cycles including functional consciousness. Importantly, we have shown how the cohesive hierarchical structure of narratives and intentions can impact an agent's processes across various time scales, thereby indirectly steering behavior. Further, rather than D, P, and M intentions being discrete items in the mind we have shown how we can coherently implement these into autonomous agents as processes. Long term coherent agential action is an achievement that emerges from the distribution of intention related processes across various time scales. Attention, self-understanding, self-narration, conscious learning, affect, memory, perception, and other processes are all, in one way or another, involved in creating the structure of agential action. If we want to understand coherent agential action, we cannot create explanations that are divorced from the functionality of consciousness. Coherent agential action is deeply integrated into the cognitive cycle, including consciousness.

Implementing agential action into a cognitive architecture in this fashion allows for future research possibilities. For example, we believe that this agential model may be able to explain philosophical puzzles regarding indecision, akrasia (weakness of will), free will, and provide insight into some of the action related symptoms of psychopathologies such as depression. We suspect that we can model psychopathologies such as depression by looking at the interplay between narratives, distal intentions, incentive salience, affect, and how such processes might go awry. Finally, understanding the relationship between narratives, distal intentions, and daily actions, might provide further input into the growing discussions regarding AI ethics [Madl & Franklin, 2015; Wallach *et al.*, 2011].

References

- Anscombe, G. E. M. [1979] *Intention* (Nachdr. der 2. ed. 1963) (Cornell Univ. Press, Ithaca, NY).
- Baars, B. [1988] *A Cognitive Theory of Consciousness* (Cambridge University Press, New York), <http://catdir.loc.gov/catdir/toc/cam032/87020923.html>.

Baars, B., Franklin, S., and Ramamurthy, U. [2007] *How deliberate, spontaneous and unwanted memories emerge in a computational model of consciousness*, <http://cogprints.org/5851/>.

Baars, B. J. [1997] *In the Theater of Consciousness* (Oxford University Press, US), doi:10.1093/acprof:oso/9780195102659.001.1, <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195102659.001.1/acprof-9780195102659>.

Baars, B. J. [2019] *On Consciousness: Science & Subjectivity - Updated Works on Global Workspace Theory* (The Nautilus Press).

Bratman, M. [1987] *Intention, Plans, and Practical Reason* (Harvard Univ. Press, Cambridge, Mass. [u.a.]).

Bratman, M. [2004] *Faces of Intention* (1. publ., digital print.) (Cambridge Univ. Press, Cambridge [u.a.]).

Bratman, M. [2007] *Structures of Agency* (Oxford Univ. Press, Oxford [u.a.]), http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=015479248&sequence=000001&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA.

Bruner, Jerome [1991] The Narrative Construction of Reality, *Crit. Inq.*, http://www.ling.upenn.edu/wlabov/L470/Bruner_1991.pdf.

Bruner, J. [2004] Life as Narrative, *Soc. Res.* **71**(3), 691–710.

Collins, P. H. [2002] *Black Feminist Thought* (Rev. 10th anniversary ed.) (Routledge Ltd, London), doi:10.4324/9780203900055, <https://www.taylorfrancis.com/books/9781135960148>.

Compton, C. A. [2020] Co-Sexuality and Organizing: The Master Narrative of “Normal” Sexuality in the Midwestern Workplace, *J. Homosex.* **67**(7), 1013–1039, doi:10.1080/00918369.2019.1582220.

Davidson, D. [1978] *Intending* (R. Yovel, Ed.).

Davidson, D. [1985] *Essays on Actions and Events* (Reprint. with corr.) (Clarendon Pr, Oxford).

Delafield-Butt, J. T., and Trevarthen, C. [2015] The ontogenesis of narrative: from moving to meaning, *Front. Psychol.* **6**, 1157, doi:10.3389/fpsyg.2015.01157.

Dennett, D. [2017] *From Bacteria to Bach and Back: The Evolution of Minds* (Allen Lane), <http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9780141978055>.

Dennett, D. C. [1992] *The Self as a Center of Narrative Gravity*, https://www.openaire.eu/search/publication?articleId=od_____68::b594448292d92ebe30a38ced4f2997c6.

Dings, R. [2019] The dynamic and recursive interplay of embodiment and narrative identity, *Philos. Psychol.* **32**(2), 186–210, doi:10.1080/09515089.2018.1548698.

- Dong, D., and Franklin, S. [2015] A New Action Execution Module for the Learning Intelligent Distribution Agent (LIDA): The Sensory Motor System, *Cogn. Comput.* **7**(5), 552–568, doi:10.1007/s12559-015-9322-3.
- Franklin, S. [1995] *Artificial Minds* (1st ed.) (MIT Press, Cambridge, Mass. [u.a.]).
- Franklin, S. [1997] Autonomous Agents as Embodied AI, *Cybern. Syst.* **28**(6), 499–520, doi:10.1080/019697297126029.
- Franklin, S. [2000] Deliberation and Voluntary Action in ‘Conscious’ Software Agents, *Neural Netw. World* **10**, 505–521.
- Franklin, S., and Baars, B. [2010] *Two Varieties of Unconscious Processes* (H. A. Collerton and F. LeBeau, Eds.; pp. 91–102).
- Franklin, S., and Graesser, A. [1997] *Is it an agent, or just a program?: a taxonomy for autonomous agents Intelligent Agents III Agent Theories, Architectures, and Languages*, 21–35, doi:10.1007/BFb0013570.
- Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., Snaider, J., Agrawal, P., and Chen, S. [2016] A LIDA cognitive model tutorial, *Biol. Inspired Cogn. Archit.* **16**, 105–130, doi:10.1016/j.bica.2016.04.003.
- Franklin, S., Strain, S., McCall, R., and Baars, B. [2013] Conceptual Commitments of the LIDA Model of Cognition, *J. Artif. Gen. Intell.* **4**(2), 1–22, doi:10.2478/jagi-2013-0002.
- Freeman, M. [2007] *Chapter 5 Autobiographical Understanding and Narrative Inquiry* (p. 120), doi:10.4135/9781452226552.n5, <https://dx.doi.org/10.4135/9781452226552.n5>.
- Gallagher, S. [2013] A pattern theory of self, *Front. Hum. Neurosci.* **7**, 1–7, doi:10.3389.
- Gallagher, S., and Hutto, D. [2008] *Understanding Others Through Primary Interaction and Narrative Practice* (pp. 17–38).
- Goldman, A. I. [1970] *A Theory of Human Action* (Prentice-Hall, Englewood Cliffs, NJ).
- Grice, H. P. [1972] *Intention and Uncertainty* (Vol. 1971) (Oxford Univ. Pr, London).
- Grysmen, A., Fivush, R., Merrill, N., and Graci, M. [2016] The influence of gender and gender typicality on autobiographical memory across event types and age groups, *Mem. Cognit.* **44**(6), 856–868, doi:10.3758/s13421-016-0610-2.
- Hofstadter, D., and Mitchell, M. [1995] *The Copycat Project: A model of mental fluidity and Analogy-Making* (K. J. Holyoak and J. Barnden, Eds.; pp. 205–267).
- Hutto, D. [2008] *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons* (Kindle) (MIT Press, Massachusetts).
- James, W. [2007] *The Principles of Psychology* (Cosimo Inc, New York).
- Kugele, S., and Franklin, S. [forthcoming] Learning in LIDA, *Cogn. Syst. Res.*
- Lae, L. L. [2019] *Culture, Self-Narratives and Autobiographical Memory: Using a Semiotic Narrative Approach to Investigate Cross-cultural Differences* [PhD Thesis], <https://trove.nla.gov.au/work/237764065>.
- MacIntyre, A. C. [2013] *After Virtue* (1. paperback ed.) (Bloomsbury, London [u.a.]), <http://bvbr.bib->

bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=028095075&sequence=000002&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA.

Madl, T., Baars, B. J., and Franklin, S. [2011] The Timing of the Cognitive Cycle, *PLoS One* **6**(4), e14803, doi:10.1371/journal.pone.0014803.

Madl, T., and Franklin, S. [2015] *A Construction Manual for Robots' Ethical Systems* (1st ed.) (Springer, Cham), doi:10.1007/978-3-319-21548-8, [https://ebookcentral.proquest.com/lib/\[SITE_ID\]/detail.action?docID=4179371](https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=4179371).

Madl, T., Franklin, S., Snider, J., and Faghihi, U. [2015] *Continuity and the Flow of Time: A Cognitive Science Perspective* (B. Mölder, V. Arstila, and P. Øhrstrøm, Eds.).

McAdams, D. [2006] *The Redemptive Self: Stories Americans Live By* (Oxford University Press, GB), doi:10.1093/acprof:oso/9780195176933.001.0001, <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195176933.001.0001/acprof-9780195176933>.

McCall, R. J., Franklin, S., Faghihi, U., Snider, J., and Kugele, S. [2020] Artificial Motivation for Cognitive Software Agents, *J. Artif. Gen. Intell.* **11**(1), 38–69, doi:10.2478/jagi-2020-0002.

McLean, K. C., Boggs, S., Haraldsson, K., Lowe, A., Fordham, C., Byers, S., and Syed, M. [2020] Personal identity development in cultural context: The socialization of master narratives about the gendered life course, *Int. J. Behav. Dev.* **44**(2), 116–126, doi:10.1177/0165025419854150.

Mele, A. R. [1992] *Springs of Action* (2. print.) (Oxford Univ. Press, New York).

Mylopoulos, M., and Pacherie, E. [2018a] Intentions: The dynamic hierarchical model revisited, *Wiley Interdiscip. Rev. Cogn. Sci.* **10**(2), e1481-n/a, doi:10.1002/wcs.1481.

Mylopoulos, M., and Pacherie, E. [2018b] Intentions: The dynamic hierarchical model revisited, *Wiley Interdiscip. Rev. Cogn. Sci.* **10**(2), e1481-n/a, doi:10.1002/wcs.1481.

Nelson, K., and Fivush, R. [2020] The Development of Autobiographical Memory, Autobiographical Narratives, and Autobiographical Consciousness, *Psychol. Rep.* **123**(1), 71–96, doi:10.1177/0033294119852574.

Pacherie, E. [2006] *Towards a Dynamic Theory of Intentions*, https://www.openaire.eu/search/publication?articleId=dedup_wf_001::2523a185eb25e4023d7865f7d01b61c2.

Pacherie, E. [2008] The phenomenology of action: A conceptual framework, *Cognition* **107**(1), 179–217, doi:10.1016/j.cognition.2007.09.003.

Ryan, K., Agrawal, P., and Franklin, S. [2019] The pattern theory of self in artificial general intelligence: A theoretical framework for modeling self in biologically inspired cognitive architectures, *Cogn. Syst. Res.*, doi:10.1016/j.cogsys.2019.09.018, <http://dx.doi.org/10.1016/j.cogsys.2019.09.018>.

Schechtman, M. [1996] *The Constitution of Selves* (1. publ.) (Cornell Univ. Press, Ithaca [u.a.]).

Schechtman, M. [2014] *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life* (Oxford university press, Oxford),

doi:10.1093/acprof:oso/9780199684878.001.0001,
<http://hdl.handle.net/2078/ebook:68127>.

Shin, Y. K., Proctor, R. W., and Capaldi, E. J. [2010] A review of contemporary ideomotor theory, *Psychol. Bull.* **136**(6), 943–974, doi:10.1037/a0020541.

Snaider, J., McCall, R., and Franklin, S. [2012] Time production and representation in a conceptual and computational cognitive model, *Cogn. Syst. Res.* **13**(1), 59–71, doi:10.1016/j.cogsys.2010.10.004.

Wallach, W., Allen, C., and Franklin, S. [2011] Consciousness and Ethics: Artificially Conscious Moral Agents, *Int. J. Mach. Conscious.* **3**(1), 177–192, doi:10.1142/S1793843011000674.