RESEARCH ARTICLE

Locating Values in the Space of Possibilities

Sara Aronowitz[1]

[1]University of Toronto
*Corresponding author: Sara Aronowitz; s.aronowitz@utoronto.ca

Abstract

Where do values live in thought? A straightforward answer is that we (or our brains) make decisions using explicit value representations which are our values. Recent work applying reinforcement learning to decision-making and planning suggests that more specifically, we may represent both the instrumental expected value of actions as well as the intrinsic reward of outcomes. In this paper, I argue that identifying value with either of these representations is incomplete. For agents such as humans and other animals, there is another place where reward can be located in thought: the division of the space of possibilities or 'state space'.

1. Introduction

Imagine a person, Masha, who grew up in a religious Jewish household and had sincere commitments about living in accordance with the religious law. Later, she changed her mind and became an atheist, explicitly disavowing the significance of rules such as the separation of dairy and meat. Encountering her right now, we might expect her explicit values to be in line with a secular life: not only her conscious thoughts or explicit speech, but even below the level of consciousness we might imagine her thoughts and feelings to be in line with her new beliefs. But at the same time, when planning her day, imagine we notice that she always makes a division between Shabbat and the rest of the week, divides days by sunset rather than midnight, and when planning what to eat for lunch will ask herself questions such as: "should I have meat or dairy?". That is, one aspect of what philosophers might think of as Masha's "descriptive" thought about the world is permeated with the older, disavowed values.

What does Masha *really* value? You might think the answer is obvious: she values a secular life. After all, that is reflected in her conscious thoughts, her emotions, and her speech, as well as her choices. In this paper, I'll try to convince you that while this is true, she *also* still values aspects of her former religious life. This is because these values survive in her view of the world, and in particular, in her division of the giant and complex present, past, and future into coarse-grained options and outcomes. These values influence her actions as well, such as when raising the question "meat or dairy?" leads to a different dinner plan than the question "spicy or non-spicy?".

As we'll see, major life changes are not the only source of a divergence between explicit and latent values. The divisions of possibility space in humans seem to have many sources. For instance, we tend to divide possibilities according to fixed properties

such as the segmentation of events by geographical and architectural features. These generic heuristics combine with event-specific features such as subjective uncertainty and the presence or absence of key people or objects. Divergences can also arise based on belief, rather than value, changes, errors, and bias. All of these factors create a situation where latent values in the possibility space are different than explicit values in a way that is systematic and interpretable.

To put the question another way. Many models of rational decision-making, both informal and formal, think of choices as reflecting two kinds of considerations: a sense of how the world works, and the way we want (or should want) things to be. For now, let "values" stand in for whichever evaluative weight in decision-making you'd like. The question of this paper is: to what extent can there be a notion of caring that is in principle separate from the way the agent partitions possibilities? The answer I'll defend is that the best accounts of caring for creatures like us cannot be this simple. While we may have a use for a narrower notion of caring that only depends on active valuing, it may be that no notion of the total evaluative shape of human decision-making can be both comprehensive enough to explain rational action and at the same time strictly independent of the division of possibilities. This is a surprising result to the extent that views that separate evaluative and descriptive outlooks are widespread in formal (e.g. Savage (1972)) and informal (e.g. Hubin (2001)) contexts, particularly in the Humean tradition (e.g. Railton (2006)). More broadly, as I'll attempt to demonstrate, the idea that values can be found in the way we divide possibilities has implications for two questions that arise for any account of decision-making: what does it mean to divide up possibilities well, and how can values be learned from experience.

In section 2.1, I first describe a reinforcement learning setting that makes the notion of a partition of possibilities, or state space, precise. Then, I define and motivate my foil: the view that values should be identified closely with features of reward systems (3). Then, I present an argument schema (4) and explore (5) how in humans, this abstract and idealized derivation is neither actual nor optimal. Instead, several independent sources feed into the process of building and modifying state spaces. In 6, I show that in principle, state spaces can be derived from explicit values along with probabilistic expectations (given certain background assumptions) using a value-of-representation approach. However, I argue, this pattern of derivation relies on assumptions about the etiology and significance of the state space that are not met in the case of human cognition. I then conclude in 7 that in us, state spaces are part of the cognitive basis of valuing, not merely a reflection of valuing. I conclude with some implications for a theory of instrumental rationality.

2. Defining the Setting

This section sets out two pieces of crucial background for what follows: the basics of a Reinforcement Learning (RL) framework, and the notion of a state space. The state space, a representation used in decision-making and planning that divides the world and the agent's actions into coarse-grained states and options, is a crucial part of my positive view. This is not a new concept, but has been to an extent obscured as part of the backdrop to philosophical decision theory rather than given direct attention. Recent interest in optimal state spaces in cognitive science (Correa et al. (2023), Radulescu and Niv (2019), Ho et al. (2022)) makes it a timely moment to bring spate spaces into

the philosophical spotlight. This is in addition to the status of RL as the most plausible current model of the cognitive underpinnings of motivated behavior.

A brief primer on the basics of RL that will be relevant for present purposes follows for readers who are unfamiliar – others may skip to 2.2.

2.1. Reinforcement Learning

Reinforcement Learning is an umbrella term for approaches to learning that bootstrap towards an optimal policy in a way that loops acting and planning. As will be seen, RL is a particularly useful framework for our question because it allows us to talk about reward which is not necessarily determined over the same state representation in which a decision takes place in contrast to classical decision theory. That is, rewards are given as a scalar signal indexed only to a time, which allows them to have very little observable structure, whereas utility functions and preferences are assigned over outcomes, which are either themselves the objects of beliefs (Jeffrey (1990)) or at least closely related to the objects of belief (Savage (1972)). This means that reward can be conceptualized in a way that is in principle separable from any particular cognitive representation of reality – and we can in turn ask how much reward itself, independent of state space representations, determines the agent's values. Further, as an influential model of human decision-making which has already been incorporating into philosophical theorizing about cognition (see Haas (2022) for an overview), RL is a good fit for bridging rational choice and human cognition.

RL is typically defined in a Markov Decision Process (MDP)[1]. An MDP is a system where the environment is in a state $s \in S$, and the agent chooses an act $a \in A$, and can receive reward R_t. The environment transitions between states based on a transition probability based on the current state and the chosen action: $P(s'|s, a)$. This setting has the Markov property, meaning the state transition function is determined by the current state and act, and no past states or acts give any more information on top of the current ones. For instance, in the game of tic-tac-toe, the states could be game boards, the acts the choices available at each turn to the 'O' player, and the state transition function the new position of the board after the move of the 'X' player. The rewards would be a point given at the end for the 'O' player winning, nothing for a draw, and a negative point for losing.

The agent is in charge of choosing how to act in each state (forming a policy π). Reinforcement learning is a family of methods for making a policy which are updated as the agent interacts with the environment, and in many conditions converge to the optimal policy, i.e. the one that maximizes (discounted, cumulative) reward (Sutton and Barto (2018)).

Here is a basic RL algorithm, Q-learning, which involves calculating a Q-value for each act-state pair according to the following formula:

$$Q_{new}(S_t, A_t) = Q_{old}(S_t, A_t) + \alpha(R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q_{old}(S_t, A_t)) \qquad (1)$$

Here α is a learning rate, and γ is a discount factor. The formula tells us how to update the Q-value for a state, act pair based on the old Q value, the reward received

---

[1]In some applications, it's more useful to relax this assumption, for instance to a partially observable Markov decision process (POMDP)

after taking the initial action and the Q value of the best action in the next state. The Q learning formula is combined with a decision rule which tells the agent how to use the Q-values to select actions. A basic rule is ε-greedy, where the highest Q-value act is selected with $1 - \varepsilon$ probability, and a random act is selected the rest of the time. The policy π is defined as mapping of states onto actions that the agent will select.

From this basic foundation, different versions of reinforcement learning can be articulated. On-policy methods like SARSA update values based on chosen acts, whereas the Q-learning equation above updates based on the highest valued subsequent act (these come apart when the agent picks an act at random). Model-based methods use knowledge of the environment to perform updates outside of the act-update loop described above, for instance by updating an entire future trajectory based on new information before acting.

For present purposes, the following features of RL are most significant. Reward is a quantity that is indexed to a time, and reflects what is of value in the environment such as winning the game of tic-tac-toe. Q-values are approximations of expected cumulative discounted reward that improve and in some cases provably converge to optimality. Agents update Q-values during action, and so do not need full information in order to plan: the updates that would have been intractable are split into smaller steps that involve (in the basic version) a single update per timestep.

2.2. State Spaces

In this section, I'll characterize state spaces and situate them within philosophical debates about choice. Classic problems of decision theory, such as Newcomb's Problem, come with a pre-defined set of options. Before we begin thinking about the agent, we already have on the table a set of options: one box or two box, stay or switch, bet A or bet B. An option set, such as {one box, two box}, is a partition over acts, and it typically corresponds closely to another partition over possible states of the world, in this case {\$1 in the opaque box, \$1000 in the opaque box}[2]. In cases like the Newcomb problem, it is natural to think about the set of options and states as given by the environment, since we don't expect well-informed agents to think of the options differently.

However, many everyday problems we face do not have such an obvious structure. On the other end of the spectrum from a well-defined set of bets dependent on well-delineated states of the world, we have truly open-ended decisions like planning what to do tomorrow. When I am deciding what to do tomorrow, there are a wide range of available ways to lay out my options and the corresponding states. I could ask myself: should I get some work done or relax? In which case I might consider as relevant questions about states: am I too tired to write well? Will my friend be around for a soccer game? And so on. Further, I might start with states and then move to acts, for instance wondering whether it will rain or not and then thinking of options by how they fit with either scenario. These options and states form the landscape of deliberation: Levi (1990) makes the case that we need a notion of options that are on the table that does not reduce to choice-worthiness, since otherwise rational deliberation could not involve a narrowing of the options leading to choice.

---

[2]The question of what this correspondence amounts to is beyond the scope of this paper.

Given this openness, we can now see why a state space can often be fruitfully considered part of the agent rather than the environment. It is up to me whether I ask myself "should I work or relax?" or instead "should I stay home or go somewhere?". Further, this does not reduce to a notion of which distinctions I can draw or which concepts I have – while I cannot ask myself "should I play jai alai?" if I lack the concept of jai alai, I can easily fail to ask myself that question even when I have the concept. It seems that truly open-ended planning is typically a problem of too many possible questions rather than too few.

But many problems we face fall in between regimented bets and fully open planning. In deciding what to eat in a restaurant, the menu gives me a partition of acts but I might consider my own partition of states, such as: is the chef Filipino or are the vegetables pre-frozen? In approaching bets in a casino, I might merge multiple gambles into one or classify outcomes by where they come in a sequence as pay-back or karma. In this sense, a 'normal' decision has some open-ended elements, and some of the variation in human behavior can be explained by how different people employ state spaces.

To return to the RL setting. In an MDP, we take a chunk of the real world and break it into abstract, typically discrete units, i.e states S and acts A. The state transition function P can only be assigned over this partition, and likewise $Q(S_t, A_t)$ assigns values to acts in states defined by the state space. Open-endedness occurs in the translation of reality into MDP, and therefore in open-ended problems, different agents face different MDPs which may have substantially different optimal strategies. In some applications, the designer of the agent makes this choice, while in others, it is the agent herself.

Imagine a match played between two players who are friends with a long history of games, but where only one has played tic-tac-toe before. We can describe this situation as I did in the previous section just in terms of board position, and idealize the transition function as depending only on the current move and position. Even in tic-tac-toe, defining the state space is not obvious. In this case, a better model of the game might depend on the track records of the players, since players tends to be riskier when they are winning overall. Then the experienced player might want to think of the states not as board positions, but as board positions plus overall score. Since the novice player is still picking up on the rules, she might rely on perceptual cues and social signals such as a slight wince as her opponent notices a mistake she's made. In this case the state space of the same match could be a rich three-dimensional scene linked to inferred mental states. This would explain why her choices depend on facial expressions and even tricks of lighting, unlike her opponent who only considers track record and board position. Even when the game is the same, different agents will approach it with a different state space, leading to different patterns of actions even when everything else about their mental states is the same.

In the case of Masha, which we started with, the questions she asks herself in planning, such as "should I have meat or dairy?" articulate a state space. Just like the division in the case of tic-tac-toe, her planning about what to eat involves assigning values to acts in states, and forming a policy of what to choose when. Had she instead asked herself "should I have spicy food?", she would be taking the same real-world problem of what to eat given her resources and tastes and modeling it according to a different partition.

In short. Which state space an agent uses determines which problem she is solving. Ordinarily, we need to know which problem she is solving to tell what she should do or to predict what she will do on the assumption that she is rational. Or in the formal

case, we need to fit an MDP before solving it. The exception to this is when there is only one problem she should be solving, in which case we can either assume that's what she's doing and determine optimal behavior accordingly or subject her to a quite different form of criticism about not seeing the problem the way she should. Thus state space choice is a key ingredient of fitting rational models in cases with significant open-endedness, which are arguably the majority of our choices.

What about in ideal agents? Jeffrey (1990)'s decision theory is famously *partition invariant*: by setting up decision problems over propositions, rather than the distinct types of acts and states, and calculating desirability of a coarser proposition as a sum over desireability of the finest grain propositions under the coarser one, the desireability of an act is never altered by changing the way the space is divided[3]. Partition invariance does not hold in Savage (1972)'s theory, for instance, so one might argue that state space choice may or may not be relevant in a complete ideal theory.[4] As we'll see, the importance of state space representations in valuing depends on features of the agent which are related to computational limitations, and so my argument does not apply to ideal agents.

Another issue with situating state spaces in the philosophical literature comes from an old debate about psychological realism. From a behaviorist perspective, state spaces are just a useful device to organize and predict the actions of agents. This could be said of any piece of the decision-theoretic model of the agent. But state spaces face an additional challenge. Thoma (2021) brings out how revealed preference theory depends on taking for granted what the agent identifies as their options: in other words, we must be able to assume a particular state space in order to define the agent's preferences in this behaviorist fashion. This, in turn, suggests that the open-ended choices in which state spaces are most usefully understood as operating is outside the purview of behaviorist theories. More generally, while we need not assume there is no way to provide a behaviorist characterization of state spaces, the most straightforward way to understand a state space is in realist terms, as the structure of the representations used in decision-making. In 5, I present some psychological models of state spaces, which gives a fuller picture of which behaviors lead to state-space ascription.

I have been referring to the state space as if there is a single one used in each choice. This is a simplification I'll use throughout this paper, which already seeks to complicate the classical picture of motivated choice. But it is worth noting that we likely use a set of different state spaces representations even in a single choice, for instance by shifting the question I ask myself when I get stuck about what to do. Models of choice in psychology often utilize structured state spaces, such as a hierarchical representation that joins state spaces at different levels of abstraction. Further complexity is added by considering that we sometimes reuse state spaces across problems and other times build ad hoc representations for each problem.

A state space is the backbone of representations used in planning and choice. It is a partition of actions and states into units that are coarse-grained enough to enable efficient and general calculation. State spaces are implied in the structure of explicit questions

---

[3]see also Joyce (2000) for a related discussion

[4]Other theorists follow Jeffrey in preserving partition invariance: Joyce (1999) preserves invariance, as does Greaves (2013) in the epistemic context, and see Arntzenius (2008) for a discussion of the link between causal decision theory and partition non-invariance.

like "should I go to the park or the beach?" but reflect features of cognition that may be pre-linguistic or implicit. While in classic choice problems, a state space is assumed based on the environment alone, in a wide range of ordinary choices we are responsible for selecting our own state space. In order to argue that state space representations are a part of (human) valuing, I will now turn to setting up a simpler model of valuing based on RL.

3. A Simple RL Theory of Value

The success of RL in artificial contexts (Mnih et al. (2015)) and as a model of neural computation (Dayan and Niv (2008)) make it a plausible model of human-like valuing. The most straightforward version of this model would be to equate reward and value. The purpose of this section is to sketch out what this identification would mean, before I go on to argue that it is incomplete. To do so will require mapping philosophical concepts onto their best operationalization in the computational model.

When considering philosophical concepts that capture the motivational contribution to decision-making, the landscape can be divided by which concepts are primary. Returning to our example, imagine Masha makes a dinner of fish and noodles for her sister. She could be said to be motivated by her desire to cook that dinner, her valuing of tasty food or making her sister happy, her valuing of her sister, or the utility of the state of her sister being happy. That is, we can describe her as related to an action (a desire to act), an outcome (utility placed on a state of the world), or an object (valuing a person). This distinction is somewhat different from the value/desire/utility distinction, since we can value or desire acts, states or objects, though utility as a more formal construct is only assigned in the first instance to states, and via states, to acts which bring those states about. For simplicity, then, I will use the term "value" in what follows, though in principle "desire" would do just as well[5].

We can now divide the space of philosophical theories as follows. Kantian theories of motivation (such as the one proposed by Tenenbaum (2010) take the primary relation to be valuing objects. In most cases, this will be a rather specific class of objects, namely persons. Outcome theories of motivation, such as realist versions of expected utility theory, take the primary relation to be valuing outcomes. Action theories, such as Small (2012), take the relation to be valuing actions. Finally, pluralist theories take there to be no single primary relation but allow for two or more relations to be equally fundamental. I will not aim to decide between these theories, but in what follows will refer to values as directed towards objects. The object approach is in general the coarsest in granularity, since a single state of valuing my sister would be translating into a variety of states of valuing actions or outcomes linked to her well-being. This coarse-grained approach, as we'll see, will allow a greater parsimony in discussing the cases under discussion.

So where are values to be found in the reinforcement learning framework? Keeping in mind that reward, R_t, is a signal given at a particular time, we might start as follows:

---

[5]A wrinkle here is that values tend to be associated with Kantianism and desires with Humeanism. In that debate, the question at issue is whether our actual motivations or some idealized version of them is more relevant to how we should act. I'll take up this issue briefly at the end of the section – it is related to the current project but extends a bit too far into the normative domain to be able to be done justice here.

- **Simple Reward Theory** To value X is to experience reward during presence or participation in states of the world linked to X's existence, well-being, or other core features.

Because the reward signal is only indexed to time, we thus would need to appeal to something else to connect reward to the object of value. In the Simple Reward theory, this connection is done by the external environment: the fact of the agent being close in time and space to the states of interest. Further, we might want the Q values, rather than the reward signal, to represent values: this would loosely correspond to the distinction between intrinsic and instrumental value, since actions like ordering an ice cream have high Q-value (since they put us on the path to getting reward) but low reward (since we have to wait a bit for the actual benefit of eating the ice cream.

More sophisticated versions of the identification of value with quantities from reinforcement learning have been proposed by Haas (*forthcoming*)[6] and Schroeder (2004) but the Simple Reward Theory will serve as a placeholder for our purposes. It will stand in for any theory of valuing in cognition that ties value exclusively to explicit, active quantities in an RL system. This explicit, RL based, view of valuing will be my foil for the arguments in the next section.

4. Argument schema

In what follows, I will argue that the explicit theory of value is incomplete, and that state spaces make a constitutive contribution to value in humans and other bounded agents. First, in 5, I will argue for an empirical claim:

- **Informational Divergence** State space representations in humans provide information about values that cannot be derived from current explicit, reward-based cognition.

This result allows us to ask: is this value information significant and interpretable? I then argue for:

- **Causal Divergence** The latent value information in state spaces is not random, but has intelligible causal connections with environmental and historical factors

These two divergence theses describe a philosophically significant phenomenon: latent value information in state spaces that is not just an accident or mistake. I then consider two ways of explaining divergence.

The first is a conservative approach, consistent with the explicit reward theory, that holds that state spaces can be evidentially connected to value but are not themselves part of valuing:

- **Mere Reflection** Latent value information in state space representations is a mere reflection of real values in explicit reward cognition.

---

[6]Importantly, Haas wants to allow for a wide time course of valuation, from standard direct attribution of subjective value, to ad-hoc valuation, and attribution in retrospect. As we'll see, Haas' view presents an interesting counterpart to my proposal: where I aim to argue that values extend farther than valuation, she instead takes valuation to extend farther than previously thought. In this sense, we end up with the result that more of cognition is value-laden than on the standard model, but by two distinct paths.

This information can tell us about the agent's history but cannot contribute to value itself.

I demonstrate that the mere reflection thesis cannot provide a satisfactory accommodation of the two divergence theses and so should be rejected as applied to humans, but can prove acceptable for a certain class of artificial agents. Then, I argue that the best explanation of divergence is:

- **Partial Constitution** Latent value information in state space representations is part of the cognitive basis of valuing.

This thesis, in contrast to mere reflection, holds of humans and a class of other agents where divergence can also be observed.

5. Divergence between Latent and Explicit Values

A rich source of data about human state spaces can be found in the psychological literature on event segmentation. Perhaps starting with Newtson (1973), event segmentation is the way in which we break up perceived events into pieces at various levels of grain. These boundaries have a significant effect on memory (Speer et al. (2007)), where dynamics such as cuing are more effective within rather than across event boundaries even when duration is held constant (Kurby and Zacks (2008)). Event boundaries reflect a variety of sources ranging from physical features of the environment (the conversation in the house vs on the beach) to subjective aspects like surprise (before Mona dropped the glass vs after she dropped the glass). There is typically a high degree of agreement across people about where to divide events, in some cases even with participants from different cultures (Swallow and Wang (2020)).

However, some elements of event segmentation do seem to vary based on background. Gerwien and von Stutterheim (2018) looked at differences in event segmentation between participants who were native speakers of French or German. In French, a new motion verb tends to be used when the moving object turns or changes orientation, whereas in German, a single verb tends to describe these sequences. For instance, in French you might say "Anna entre et monte les escaliers" (Anna comes in and goes up the stairs) whereas in German "Anna kommt herein und die Treppe hinauf" (Anna comes in and up the stairs). Gerwien and von Stutterheim (2018) find that these differences are correlated with differences in non-linguistic segmentation, where participants watch videos of events and are asked to press a key when a new event begins: French participants were more likely to start a new event when direction of motion changes, whereas German participants did not mark these changes. Note that from a global perspective, speakers of French and German are closely related groups speaking similar languages.

We can then separate event segmentation factors into three general categories: (a) generic principles shared among all people, such as indoors/outdoors or the salience of loud noises, (b) culturally specific but consistent within a group principles such as linguistic conventions of segmenting motion, and (c) individual specific effects, many of which may arise from more general principles, such as marking a boundary when a surprising event occurs (see Fig. 1). The evidence suggests that all three categories are significant sources of human event segmentation. The example of Masha which we started with points to another division in the individual sources of segmentation in (c).
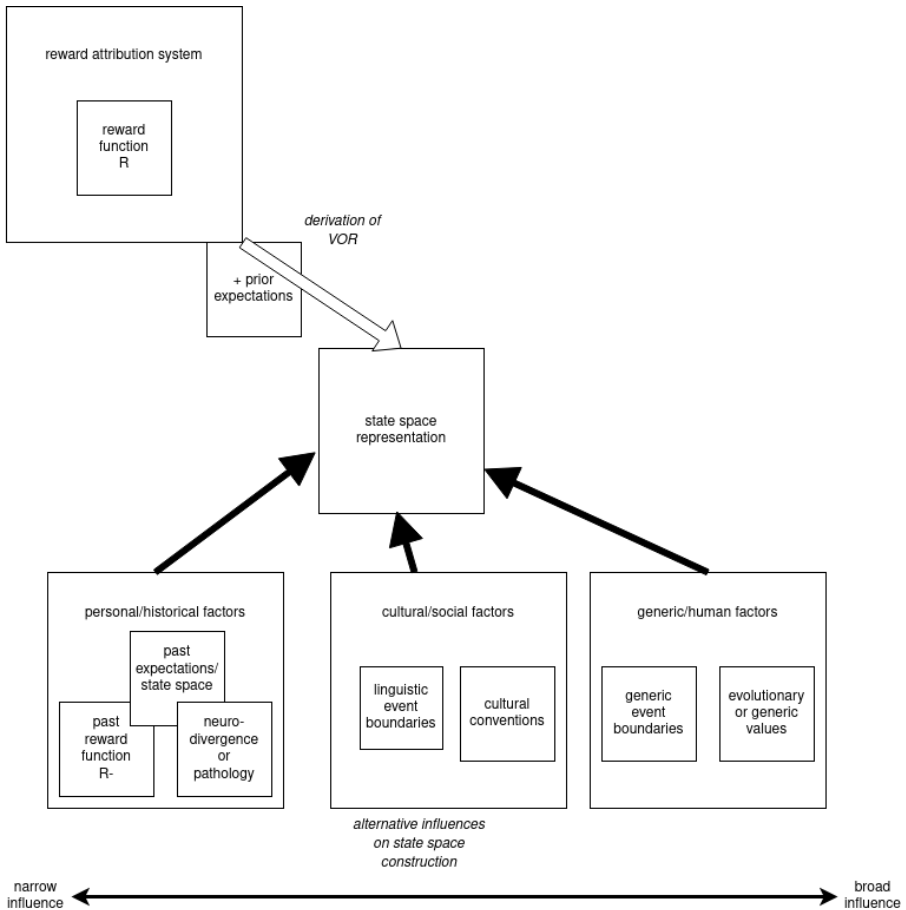
Figure 1. A schematic of influences on state space representations. Note that these nodes depict synchronically distinct lines of influence on the state space representation, but they are likely interconnected at least diachronically. The current reward function influences future personal/historical factors, and both of these can influence cultural and social factors, for instance by causing the agent to seek out a different subcultural group.

In her case, individual features from her past were the source of some of the ways she divides the state space in the present. While I can't identify any attempts to investigate this hypothesis of past values in current state spaces empirically, we might view it as a consequence of two well-supported features of human memory: (a) as already discussed, that event segmentation (or the determinants thereof) are stored in memory, and (b) that stored memories do not always quickly update when new information is presented, especially when there is no intervening retrieval event (Johnson and Seifert (1999), Schiller and Phelps (2011)).

Beyond separating a stream of occurrences into events, state space representations also involve segmenting or coarse-graining outcomes. Using an RL framework, Radulescu and Niv (2019) argue that differences in mood dynamics in bipolar and unipolar depression could be understood as downstream results of altered state space

representations in these populations. For instance, if I ask my friend if I talked too much in class, I might divide the possible answers into she says yes or no. But I might also think of the possibilities as 'she says yes and I do talk too much' and 'she says no but she's just being nice'. A tendency to interpret outcomes like this could send me into a downward spiral in mood even when someone else who received the same external input and structured the outcomes differently would not. This example is exaggerated but Radulescu and Niv (2019) present a less obvious setting where these dynamics are at play to create substantially different outcomes.

At the cultural level, qualitative work has identified group-level biases that might be thought of as involving drawing state-space boundaries. Mills (2007), discussing white ignorance, isolates a pattern of thinking and speaking that reflects a profound neglect of the situation of non-white people and the history behind it. Importantly, this neglect is both active and also primarily works by submersing relevant events and questions below the threshold of attention. Ignorance in his sense is sustained by deliberate inattention. If this is correct, part of inattention is not distinguishing possibilities and so this would be a case where we would expect some cultural groups – e.g. white Americans – to display a culturally specific pattern of state space divisions.

Taking all of this together, we have identified informational divergence at different levels in human thinking. These findings are not predicted by the hypothesis that state spaces are optimized to explicit, individual values at the time of use. That is, this evidence supports informational divergence.

These levels of informational divergence correspond to different levels of influence on human state space construction aside from current reward: individual, such as the influence of past values, cultural, such as event boundaries modulated by language, and generic, such as the preference to place a boundary between events occurring inside and outside. These are forms of causal influence. In the individual case, we see that Masha's past values caused her current division of meat and dairy in a way that is not directly mediated by her current explicit values. This must be true, given that her current explicit values conflict with both the current state space representation and her past explicit values. At the cultural level, in the case of linguistic event boundaries, a causal story might connect early language learning with scene perception, for instance, or perhaps an influence of language on semantic categorization of events. For generic divisions, we might imagine these reflect innate prior expectations, or learning of clusters of perceptual features from experience, or something else entirely.

Overall, it seems fair to suppose that given how widespread commonalities in event segmentation are, and how much variety we observe in explicit values of human agents at any particular time across the lifespan and population, these ways of dividing up the world have their origin in something beyond a connection to explicit values that is updated continuously. It could be the case that these divisions are an ecologically valid heuristic for maximizing the value of representation, but the very idea of such heuristics, such as generic inductive biases (Simon (1978); Todd and Gigerenzer (2007);Lieder and Griffiths (2020)), is inconsistent with a perfect and synchronic causal connection. The computational savings associated with such heuristics would be undermined if they were continuously brought about by the product of a calculation.

These divergent sources of state boundaries should be particularly surprising when we think about the complexity of optimizing state space representations. In terms of computational cost, maintaining coherence between the state space and reward function

as values change is expensive both in the sense that it requires constant updates to choose a new state space and in that once one is chosen, past knowledge has to be constantly translated to the new representation. In terms of risk aversion, looping effects between new values and new state spaces could lead to catastrophic error. In addition to empirical evidence, then, we can add that is not unreasonable that a creature like a human with limited capacity and changing values might evidence a gap between explicit and latent values.

To put this point another way, event boundaries are features of long-term memory. While we see some retrospective changes to boundaries (Zadbood et al. (2021)), many features of long-term memory are resistant to change. Given changing values, making retrospective changes to event boundaries to keep up with these changes would be computationally costly, risky, and (when values can revert back to former states) redundant. This argument tells a rational story about cases like Masha's, and helps to explain why we see informational divergence.

In summary, I've argued for two claims. First, that psychological research on event boundaries suggests that state spaces are informative about values above and beyond what we find in explicit synchronic rewards. Second, that this connection makes sense in light of environmental and historical circumstances which modulate state spaces in ways that are interpretable and may even be boundedly rational.

6. An Alternative Explanation: Mere Reflection

In this section, I consider, and then reject, the thesis that state spaces are related to values insofar as they merely reflect what the agent values. According to the Simple Reward theory, the state space is part of the scene on which values are located, but not part of those values. Could this be enough to accommodate divergence?

6.1. The Mere Reflection Approach

Divergence, on the Simple RL theory, is a mere reflection of value. We can make this idea precise by considering the problem of making an RL agent. Imagine we are designing an agent to maximize cumulative reward in a set of environments. For instance, we might fix rewards as escaping a maze and minimizing travel time, and the environment as a probability distribution over a set of hedge mazes. Once we have fixed this question, we can think of the best state space using a value-of-representation (VOR) approach. This means defining the utility of a representation in decision-making as the expected utility of acting on the best plan made with that representation. Ho et al. (2022) use exactly this approach to define a value of a construal c, a simplified state space representation used in planning, over which we can define a policy π_c. Each policy has a utility equivalent to the expected value of taking the acts suggested by the policy in the states specified by the construal:

$$U(\pi_c) = U(s_0) + \sum_a \pi_c(a|s_0) \sum_{s'} P(s'|s_0, a) V_{\pi_c}(s') \qquad (2)$$

This gives us a measure of how good a policy is, and we could then immediately define the value of a representation as that of the best policy that could be built using that representation. But different representations will also differ as to their computational

costs. Some are simpler, faster, and so forth than others, and so the full VOR is defined as:

$$VOR(c) = U(\pi_c) - C(c) \tag{3}$$

Intuitively, the best representation is the one that leads to taking acts with the highest expected utility given the costs of using that representation.

With this in mind, we can return to our maze-solving agent. Taking our set of hedge maze environments, which are three-dimensional and full of details, we could pick out a set of candidate state spaces that represent the environment and action types at a coarser grain. For instance, one family of state spaces would represent the maze as a 100 x 100 grid, and actions as moves in 4 directions from each cell. Another would render a 3-dimensional environment but ignore color and texture. Amongst a suitable set of these, and given a measure of computational cost, we could calculate which state spaces have the highest VOR relative to the reward function already fixed (high reward for escape, small penalty for travel time). This would be the optimal state space (or set of state spaces) for this scenario of agent, environments, and reward.

I also want to note here that we could use the exact same machinery to diagnose a mismatch between a scenario and a state space, and to more generally infer backwards from state spaces to scenario ruled in and out as optimal given those state spaces. Further, we can do triangulations such as: given a state space, environment set, and agent properties, infer consistent and inconsistent reward functions. In other words, relative to a background context, the state space encodes latent value information in the form of constraints on which reward functions would be consistent with that state space. The more information we include in the context, such as environment set, cognitive constraints, and so on, the more latent reward information in the state space.

For example, across the same set of hedge mazes, we might design one agent to cut the hedges, one to navigate the mazes, and one to move obstacles out of the paths. Given these different goals, we might fix different reward signals and then use these together with the environment set and cognitive costs to specify a state space for each agent. The agent who trims the hedges will need to take different actions but also recognize different features of the environment: it would need to represent the length and height of the hedges. The navigating agent might get by with a very simplified representation that only represents differences in 2D floor plan or graphical representation of the maze layout. The obstacle-moving agent would need to represent the location and weight of the objects in the path, along with properties relevant to navigating in and out. The idea of latent reward information is just that given just these state spaces and none of the reward information, we could already guess which state space representations belong with which goals of the agents. This won't be a one-to-one mapping by any means, but the more information we have about the environment and agent, the more constraints the state space imposes on possible reward signals which would render that state space optimal under the VOR measure.

For our hedge maze walker, then, the reward signal is prior to the latent value information. Prior both in terms of the order of hypothetical construction, and also prior in terms of information: the reward signal, plus the context, tells us everything we can know about the state space representation, since this is how it was chosen. The reverse

will not be true except in special cases[7]. In this case, then, it seems completely uncontroversial to describe the explicit reward signal as the sole place where we find the agent's values. So far, this story fits perfectly with explicit RL theories of value.

6.2. Mere Reflection isn't Enough

We previously saw two senses in which the explicit rewards are prior in the hedge maze walker: causal and informational. Other things being equal, if both of these kinds of priority fail, we might then conclude that there is no longer a good reason to think of the latent value information as merely reflecting value rather than as making a distinct contribution to value. But we need to make that more precise.

Causal priority, in the case of the hedge maze walker, was priority in the narrative of creature construction, reflecting a kind of order of necessity or rationalization. We start with X and then add Y because X is part of the structure of the problem we find ourselves with, and Y is part of the solution. In fact, the reward signal does not need to be treated as prior in this sense: Abel et al. (2021) for instance, take up the problem of starting with a desired behavior and then inferring reward structure that will help achieve those behaviors. For instance, when designing a robot vacuum, it would be a bit odd to start with a reward signal assigning value to states and times - instead, I most directly want the vacuum to perform a behavior: roam the room and collect all the dust. I can then think of reward signals as instrumental to generating these behaviors.

Suppose we concede that the reward signal is causally prior to the state space representation in the hedge maze walker. Under what conditions would the reward signal fail to be causally prior? In principle, the reward signal and state space representation might be totally independent, as is the case when I use a default state space representation such as a pixel-wise grid for representing a video game. In a biological creature, a state space might be learned from the environment in a way that does not depend on the agent's explicit reward function, such as when I divide time by the days of the week simply because it is conventional in my society. However, in any natural or artificial creature where state spaces and reward signals are mutable, we might then have a subsequent kind of causal influence between the two even if they were originally independent. I might continue to use the days of the week representation because it aligns with my interests, or alter it to do so. In this case, the explicit reward signal can still be causally prior to the state space even if they each arose independently.

Informational priority has the same structure. Rather than consider relationships between these representations at an origin point, in realistic agents we will need to consider them as they evolve through learning and other kinds of change. The reward signal is informationally prior in the case of the hedge maze agent because it is maximally informative (given a scenario) about the state space without the reverse being true. In the original case, we imagined the state space chosen was a single optimal solution but in many cases there may be a large set of equally optimal representations. In that case, informational priority is consistent with not having enough information to get every property of the state space right - instead, the reward signal gets as close as any predictor could get. This type of informational priority is still very strong. In order to

---

[7]This is because the same state space might be optimal on many different reward-expectation pairs, so a unique function cannot be inferred backward from state representations.

be charitable to the explicit reward view, then, we will consider informational priority to fail not whenever we can do better at predicting the state space given some further information, but only when this prediction is *interpretable* and the predictive value is *substantial*. This excludes cases where, for instance, I chose among equally good state spaces by using a nursery rhyme. In that case, the predictive value of knowing the nursery rhyme is marginal, and the relationship between the nursery rhyme and the chosen space is arbitrary or uninterpretable.

Now we can specify what is needed for both of these claims to fail, and thereby to undermine the rationale of treating the reward signal as merely reflected in the state space. First, examining the state space of an agent needs to give us tangible and significant constraints on what the agent must value, given a background scenario. Second, the state space cannot be derived from the reward signal and scenario either causally or informationally, either initially or through the process of development and learning. Instead, there must be another causal source for the state space, and the state space contain substantial information that can be meaningfully predicted not by the reward signal but by some other source in an interpretable way. Of course, this is precisely what we failed to identify in the case of human state spaces in 5. So I conclude that the mere reflection view does not explain widespread features of human cognition.

7. In Favor of the Latent Values Explanation

Setting aside mere reflection, how might state spaces be related to valuing in humans? The best explanation of divergence is that the latent values in an agent's state space representations are an ineliminable part of the answer of what the agent values. This is the partial constitution thesis, which is compatible with thinking that explicit rewards are also an important part of values in thought.

From a cognitive perspective, partial constitution might be a surprising view. Even if the brain is the most well-behaved kind of reinforcement learning engine, we still won't have a single neat component of this system to label as the representation of value. But this pessimistic emphasis is not the whole story.

First, the idea that our way of setting out a decision says something directly about what we value has been defended on independent grounds by Seidman (2008). He argues that decision-making over a coarse-grained partition that does not cover all of logical space is a reasonable way of making decisions for agents like us, and one that coheres with a variety of popular ethical theories. Caring about someone is in part a question of attending to them and their interests, and drawing the correct distinctions is crucial to attention.

Another reason to accept the theory that state representations are part of valuing is that it has more resources for mapping values and beliefs onto behavior. Often, state space representations are crucial for successful behavior. As a simple illustration, imagine Hamza values birds, in that when he is faced with a decision, he always picks the option that helps birds the most. He is buying windows for his apartment and divides the space of all the windows he could purchase into single and double-pane windows. As it turns out, single pane windows are on average a little worse for birds, so he chooses the double-pane and the store delivers him one of their double-pane window sets in his price range. However, a better predictor of bird harm is the coating on a window: visibly coated windows are much safer for birds than transparent coated windows. Hamza

knows this, but doesn't use it to structure his planning. Had Hamza divided the decision into coated/uncoated instead of single/double, he would have chosen a window that better achieved his goals.

This example demonstrates the well-known point that agents can have all the right expectations and values but still fail to achieve reasonable ends when the state space is mis-specified. That is, Hamza is doing as best as he can by his own standards *within the decision problem*, but would be doing much better from an external perspective had he divided up the world in a different way. For a reinforcement learning agent, the problem can even be defined in a more internal way. Recall that according to the Reward Attribution Theory, the way in which the reward system is assigning rewards fixes what the agent values. Now assume that the reward system is assigning rewards using a different partition than Hamza's decision-making utilizes. In this case, Hamza is not just doing worse than he could be according to an observer, but he is falling short of a standard derivable from his own thought, just not from the decision problem he has currently set up. While this kind of reward information is sometime described as external to the agent (Sutton and Barto (2018)), in a human being it is clearly internal to a person's thinking.

In this sense, adopting the view that latent state space information is part of valuing allows us to criticize agents like Hamza. If he consistently fails to divide the world into categories pertinent to birds, we would now describe him as ambivalent: according to his explicit values, he cares about them, but according to his latent values, he doesn't. On the explicit reward theory, he values birds fully but fails to think about the world correctly. The key difference between these descriptions is this: on the explicit theory, Hamza should revise his state space, whereas on the implicit/explicit theory, he could either revise his state space or bring his values into alignment by revising explicit values, or some combination of the two.

This flexibility is desirable for two reasons. First, when latent values reflect important concerns are brought about by pertinent evidence, and so on, it might be wrong to revise these instead of explicit values. For instance, imagine a reverse of Hamza's case: Zaid doesn't start out caring about birds but he spends a lot of time with his brother who knows all about them. Over time, Zaid starts dividing the world in the way that his brother does just by habit – he becomes used to planning with his brother, and adopts some of the ways of talking and thinking that reflect a concern for birds. On the implicit/explicit view, Zaid now has a reason to adapt his explicit desires towards his implicit ones: that it is best, all things being equal, to have a coherent set of values. This verdict isn't possible on the explicit reward view – instead, Zaid makes a mistake in adapting his state space at all.

I am not sure whether the verdict I am advocating for in Zaid's case is intuitive. But this kind of answer has theoretical value as a first step towards solving an important problem. Namely, how can we improve our values? Philosophers and others using the machinery of rational choice theory have tended to allow that values are improved only when they become more coherent, but that any coherent set of values is uncriticizable from the perspective of practical rationality. This paired with an explicit reward thesis has the result that reward functions, when coherent, always get it right.

But adding a component of implicit reward changes the calculus without doing away with the centrality of incoherence. State space construction, as we've seen, has a variety of sources, including components we might think of as descriptive such as placing

boundaries at natural divisions in scenes. These sources feed into state spaces, which are treated as their own source of value constraints. The cases of Zaid and Hamza illustrate the result: more incoherence is possible, since there are now more grounds for value that can diverge, and consequently revisions can take a more flexible form. We can revise explicit to match implicit, implicit to match explicit, or something in between. This means agents are more often criticizable for their values, and we have more resources to recommend alterations: for instance, we might appeal to the validity of the external sources of state space divisions, the relationship of both explicit and implicit values to successful behaviors, and so on.

Thus partial constitution presents a view of state spaces where these representations are not just reflections of value but repositories of value. This explains divergence, not just by showing how it would be possible, but why it might be computationally efficient: updating state spaces is costly, and changing values too fast can be costly too. Using state spaces to store value information then exploits the inertia of these representations to simplify computation and build in risk aversion. The partial constitution view also has surprising normative consequences: in blurring the boundary between descriptive and evaluative forms of cognition, we have also revealed an interface between descriptive and evaluative norms.

8. Conclusion

We have preliminary evidence to think that reinforcement learning models will be apt models of decision-making in humans and other animals. Does this mean that reward, or a related quantity,is the cognitive realization of value? The family of explicit reward theories claim that it is. These theories are powerful and linked to empirical findings, but I've argued that even if the human mind or brain is an excellent fit to reinforcement learning models, the simple answer is incomplete.

In the toy model of the hedge maze walker, state spaces, or coarse-grained partitions over possible states and acts, are derivable from and causally dependent on reward information. In this case, they are merely reflections of value, not sources of value. But in creatures like us, this relationship breaks down. Generic, cultural, and personal factors all feed in to state space choice beyond and outside of their influence on current explicit reward. This makes state spaces neither causally nor informationally secondary to explicit values. For this reason, I've argued the explicit reward theories don't fit our case, and that instead we should view the latent value information encoded in state spaces as a distinct and self-standing locus of value.

The consequences of this shift are that agents whose current explicit reward-based values and latent state space values diverge, such as all of us, are not facing a gap between faultless values and flawed representations. Instead, the partial constitution view I've defended describes these agents who have a systematic and interpretable divergence between sources of values as ambivalent. The normative recommendation for such agents might be to alter explicit values, implicit ones, or both. This allows for a novel route for value change which I would go so far to call learning: the environment alters state space representations, and these drive an update to the reward system.

We started with the character of Masha, who would on the simple view be classified as whole-heartedly adopting a new value system though suffering from a lag in her descriptive outlook on the world. On my framework, Masha is actually ambivalent: she

in a sense is indifferent to whether she mixes meat and dairy, and in a sense cares about it, and more generally in a sense has moved on from her previous concerns, and in a sense still holds on to her past values. This is not a merely semantic issue. On the standard view, Masha should update her descriptive outlook, whereas on my view, she could resolve her incoherent state by moving in either direction. Seeing Masha as ambivalent also leads us to recognize state spaces as an arena for theoretical and practical concerns to combine and conflict, and opens up a route by which merely descriptive knowledge can put rational pressure on our values.

Acknowledgements

References

Abel, D., Dabney, W., Harutyunyan, A., Ho, M. K., Littman, M., Precup, D. and Singh, S. (2021) On the expressivity of markov reward. *Advances in Neural Information Processing Systems. 34*, 7799–7812.

Arntzenius, F. (2008) No regrets, or: Edith piaf revamps decision theory. *Erkenntnis. 68*, 277–297.

Correa, C. G., Sanborn, S., Ho, M. K., Callaway, F., Daw, N. D. and Griffiths, T. L. (2023) Exploring the hierarchical structure of human plans via program generation. *arXiv preprint arXiv:2311.18644*.

Dayan, P. and Niv, Y. (2008) Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology. 18*(2), 185–196.

Gerwien, J. and von Stutterheim, C. (2018) Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition. 180*, 225–237. https://doi.org/10.1016/j.cognition.2018.07.008.

Greaves, H. (2013) Epistemic decision theory. *Mind. 122*(488), 915–952.

Haas, J. (2022) Reinforcement learning: A brief guide for philosophers of mind. *Philosophy Compass. 17*(9), e12865.

Haas, J. (forthcoming) *The evaluative mind*.

Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D. and Griffiths, T. L. (2022) People construct simplified mental representations to plan. *Nature. 606*(7912), 129–136.

Hubin, D. C. (2001) The groundless normativity of instrumental rationality. *Journal of Philosophy. 98*(9), 445–468.

Jeffrey, R. C. (1990) *The logic of decision*. University of Chicago press.

Johnson, H. M. and Seifert, C. M. (1999) Modifying mental representations: Comprehending corrections. *The construction of mental representations during reading*. pp. 303–318.

Joyce, J. M. (1999) *The foundations of causal decision theory*. Cambridge University Press.

Joyce, J. M. (2000) Why we still need the logic of decision. *Philosophy of science. 67*(S3), S1–S13.

Kurby, C. A. and Zacks, J. M. (2008) Segmentation in the perception and memory of events. *Trends in cognitive sciences. 12*(2), 72–79.

Levi, I. (1990) *Hard choices: Decision making under unresolved conflict*. Cambridge University Press.

Lieder, F. and Griffiths, T. L. (2020) Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences. 43*, e1.

Mills, C. (2007) White ignorance. *Race and epistemologies of ignorance. 247*, 26–31.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G. *et al.* (2015) Human-level control through deep reinforcement learning. *nature. 518*(7540), 529–533.

Newtson, D. (1973) Attribution and the unit of perception of ongoing behavior. *Journal of personality and social psychology. 28*(1), 28.

Radulescu, A. and Niv, Y. (2019) State representation in mental illness. *Current opinion in neurobiology. 55*, 160–166.

Railton, P. (2006) Humean theory of practical rationality In *The Oxford Handbook of Ethical Theory*, Copp, D. (eds). Oxford University Press. pp. 265–81.

Savage, L. J. (1972) *The foundations of statistics*. Courier Corporation.

Schiller, D. and Phelps, E. A. (2011) Does reconsolidation occur in humans? *Frontiers in behavioral neuroscience*. *5*, 24.

Schroeder, T. (2004) *Three faces of desire*. Oxford University Press.

Seidman, J. (2008) Caring and the boundary-driven structure of practical deliberation. *J. Ethics & Soc. Phil. 3*, i.

Simon, H. A. (1978) Rationality as process and as product of thought. *The American economic review*. *68*(2), 1–16.

Small, W. (2012) Practical knowledge and the structure of action . In *Rethinking Epistemology*. De Gruyter. pp. 133–228.

Speer, N. K., Zacks, J. M. and Reynolds, J. R. (2007) Human brain activity time-locked to narrative event boundaries. *Psychological Science*. *18*(5), 449–455. 10.1111/j.1467-9280.2007.01920.x. PMID: 17576286.

Sutton, R. S. and Barto, A. G. (2018) *Reinforcement learning: An introduction*. MIT press.

Swallow, K. M. and Wang, Q. (2020) Culture influences how people divide continuous sensory experience into events. *Cognition*. *205*, 104450. https://doi.org/10.1016/j.cognition.2020.104450.

Tenenbaum, S. (2010) *Desire, practical reason, and the good*. Oxford University Press, USA.

Thoma, J. (2021) In defence of revealed preference theory. *Economics & Philosophy*. *37*(2), 163–187.

Todd, P. M. and Gigerenzer, G. (2007) Environments that make us smart: Ecological rationality. *Current directions in psychological science*. *16*(3), 167–171.

Zadbood, A., Nastase, S. A., Chen, J., Norman, K. A. and Hasson, U. (2021) Here's the twist: how the brain updates the representations of naturalistic events as our understanding of the past changes. *bioRxiv*. pp. 2021–09.