Trends in
Cognitive Sciences

CellPress
OPEN ACCESS

Review

Goals, usefulness and abstraction in value-based choice

Benedetto De Martino [1,2,*] and Aurelio Cortese[1,2,*]

Colombian drug lord Pablo Escobar, while on the run, purportedly burned two million dollars in banknotes to keep his daughter warm. A stark reminder that, in life, circumstances and goals can quickly change, forcing us to reassess and modify our values on-the-fly. Studies in decision-making and neuroeconomics have often implicitly equated value to reward, emphasising the hedonic and automatic aspect of the value computation, while overlooking its functional (concept-like) nature. Here we outline the computational and biological principles that enable the brain to compute the usefulness of an option or action by creating abstractions that flexibly adapt to changing goals. We present different algorithmic architectures, comparing ideas from artificial intelligence (AI) and cognitive neuroscience with psychological theories and, when possible, drawing parallels.

Highlights

Value has often been used as a synonym for reward, with a focus on its hedonic aspect while overlooking its functional concept-like nature. Recent research has started to highlight its functional and goal-dependent aspects by directly manipulating the goal of the task and introducing the concept of usefulness.

Constructing value representations of usefulness involves the process of abstraction, thus reducing dimensions and coding only relevant information. Neural mixed selectivity may be the underlying coding principle of abstract value representations.

Information is selected through cognitive mechanisms like memory and attention. This process requires crosstalk between different brain regions that include sensory cortices, the hippocampus, and the prefrontal cortex.

Metacognition can provide a mechanism for the concurrent roles of monitoring and updating abstract value representations. Algorithmic architectures and their neural implementations are presented and discussed.

The flexibility of value-based choices

What is value? In everyday language, value and **reward** (see Glossary) are often used interchangeably: if something is rewarding, then it is valuable. Conversely, things are valuable because they are rewarding (e.g., I value coffee since I find drinking it very rewarding). But there is more to value than its rewarding aspect. For example, consider for a moment your dining table. You may not have thought about it much before now or thought of it as a valuable possession. But imagine that while dining, your room starts shaking, and the chandeliers begin to swing. Books, pots, and frames topple from the shelves. You are in the middle of an earthquake. Suddenly, that unassuming table might have just become the most valuable object you own.

How do you judge whether your table is indeed a valuable earthquake shelter (Figure 1)? After all, the table is not intrinsically valuable or rewarding. It acquires its value by construction as a shelter and only for pursuing a precise **goal** (protecting your body in an earthquake). Furthermore, the table's value as a shelter is determined by only a subset of its features: the sturdiness of the tabletop and the thickness of the material are relevant, but its colour is not. This small subset of relevant features must be selected from perception through attention and selective retrieval from memory. For instance, if you (similarly to the authors of this review) grew up in a region prone to earthquakes, you might remember a safety drill in your school. Or you might recall seeing on the news the images of Japanese citizens finding shelter under tables during the Kobe earthquake of 1995. But even if you lack such memories, you would still be able to generalise from what you have learnt throughout your life: hard and flat surfaces can protect your body from falling objects.

How do we accomplish these feats? To answer this question, we will first look at the difference between reward, value, and utility, providing a brief overview of how the fields of economics and psychology have differentiated these concepts. We will then present the computational principles that allow the brain to build a low-dimensional abstract value representation to fulfill a particular goal. Following this, we will dive into the cognitive machinery used to select the

[1]Institute of Cognitive Neuroscience, University College London, London WC1N 3AZ, UK
[2]Computational Neuroscience Laboratories, ATR Institute International, 619-0288 Kyoto, Japan

*Correspondence:
benedettodemartino@gmail.com (B. De Martino) and
cortese.aurelio@gmail.com (A. Cortese).

information that should be included in these abstract value representations. Finally, we will present some putative mechanisms that allow the brain to monitor, update, or replace value abstractions that stop being or fail to be useful.

Value, reward, and utility

Utility and reward in economics

The true nature of value has puzzled scholars for centuries. The English moral philosopher and social reformer Jeremy Bentham described value as 'that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness...[or] to prevent the happening of mischief, pain, evil, or unhappiness'. According to Bentham's view, value strongly resembles the definition of reward (positive value) or punishment (negative value), yet a clear division between reward and value was not made explicit in his definition. Roughly at the same time, Swiss mathematicians Nicolas and Daniel Bernoulli realised that humans' choices are not always driven by the maximisation of external objective rewards but by an internal representation of value, so-called 'utility'. Imagine that you are playing a game of chance in which you are asked to repeatedly flip a fair coin. You start from £2: every time 'heads' comes up, the amount doubles (£2 - £4 - £8 - £16 - £32 ….); however, if 'tails' comes up, you lose everything. The objective expected reward doubles on each flip *ad infinitum*:

$$\begin{aligned}&\text{Coin flip (expected value)}\\&=\frac{1}{2}\times 2+\frac{1}{4}\times 4+\frac{1}{8}\times 8+\frac{1}{16}\times 16\ldots\\&=1+1+1+1\ldots=\infty\end{aligned} \quad [1]$$

Most of us will probably decide to stop playing this game after ten lucky flips or less. What we have described is called the St. Petersburg paradox, a problem invented by Nicolas Bernoulli and analysed by his cousin, Daniel. The latter suggested that utility (U) follows a logarithmic relation:

$$U(w)=\ln(w) \quad [2]$$

where w is the total wealth of the gambler [1].

The key intuition in Bernoulli's analysis is that people's choices are not determined by a numeric objective reward but instead by an internal representation of utility. That is, subjective value or utility might be different from an objective numerical reward.

But how is this subjective utility constructed by an individual in the first place? Animal behavioural theories and the field of reinforcement learning (RL) have provided some answers.

Learning the value of options and actions: RL

First, some utilities are not constructed but have instead been sculpted in our genes by our evolutionary history. A baby does not need to learn that milk is valuable, while a loud noise is not. These innate values can be extremely useful in some circumstances, they can even save our lives, but they are limited in scope and egregiously inflexible.

A more flexible way to construct utility during repeated interaction with the environment is through **model-free learning**. The goal of the agent is to maximise the total amount of current and future rewards. Value is therefore built through iterative updates of trial and error (Box 1). For well-defined classes of problems, model-free architectures can be powerful and in their more complex

Glossary

Cognitive control: the set of processes upon which an agent selects relevant information and inhibits irrelevant information to attain a specific goal.
Context: a configuration of the environment that determines the nature of the outcome of the agent's actions over time. Can be explicit but also implicit (i.e., decision-maker might be unaware or not access it verbally). It is usually distinguished from cues and stimuli.
Cortico-thalamic loops: the set of connections linking cortical areas such as occipital, parietal, or prefrontal cortices to the thalamus, a set of nuclei found in the inner, subcortical part of the brain.
Curse of dimensionality: the problem arising in learning when the dimensionality of the space is too large for the algorithm to converge to a solution by brute force.
Dimensionality reduction: transformation of data from a high-dimensional space to low-dimensional space, while retaining core properties of the original data.
Goals: the object of an agent's effort. A goal is often endogenous and is characterised by intentionality (i.e., sense of agency) with (usually) an active component from the decision-maker.
Mixed selectivity: the property harboured by many neurons in prefrontal regions; display complex responses tuned to multiple stimuli, features, context, or combinations thereof.
Model-based learning: in RL, a class of learning algorithms in which the agent builds a predictive model of the environment (i.e., transition and reward functions) based on its interaction with it.
Model-free learning: in reinforcement learning, model-free algorithms learn the consequence of an action by adjusting their strategy (i.e., policy) to maximise overall reward, but they do not require a representation of the dynamic of the environment (i.e., state transitions and reward functions).
Reward: (hedonic value); construct used in neuroscience, psychology, and economics to describe the pleasurable aspects of a positive outcome.
Rule(s): similar to context, it is a specific mapping of environmental variables and agent's action/response with outcomes. However, unlike context, it is always explicit (i.e., known to the decision-maker and verbally accessible). It is often not set by the agent and is limited to specific circumstances or amount of time.

instantiations, such as deep RL, can achieve superhuman performance [2–4], but are otherwise slow when goals change, resulting in less flexible behaviour.

The most flexible system for assigning utility to the environment or one's actions is **model-based learning**. At the computational level, model-based RL requires a more complete understanding of the environment structure [5–8]. Interactions between midbrain dopamine areas and the striatum are thought to provide the substrate for reward predictions (value), which could largely influence goal-directed processes. However, there are constraints on the computational capability of the brain and how humans learn from small training samples [9] that are not yet fully understood.

These learning algorithms must contend with the fact that the information we receive from the environment is rich and complex. Therefore, the brain must construct a compact description of the current setting that is relevant to the current goal: a 'state', as commonly referred to in RL. Think about the table example and imagine you are a small child who is learning which surfaces are good for stopping falling objects. The colour of the surface should not be part of your state, but its material or thickness should. However, if your goal changes and you want to determine whether that same table is a good surface on which to play marbles, the thickness would be irrelevant while the smoothness and the colour (since this might make the marbles more visible)

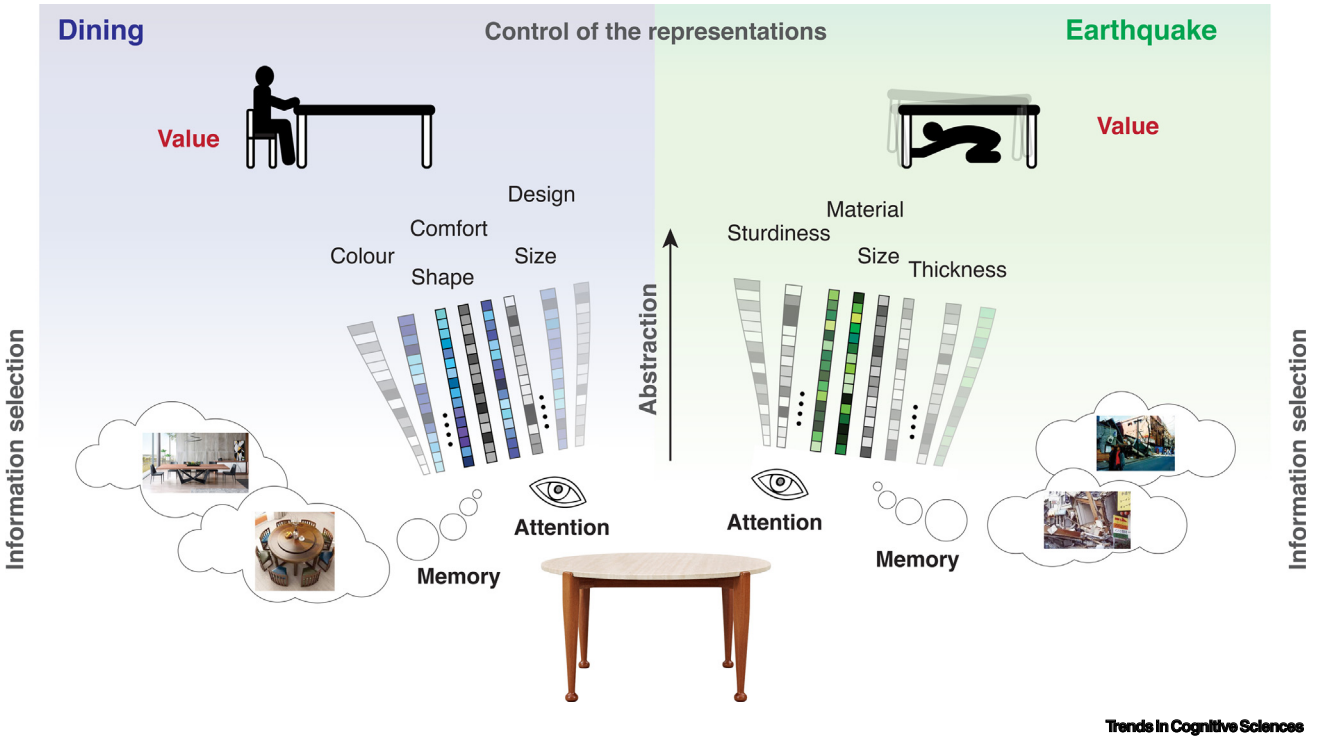

Figure 1. The construction of abstract, goal-based value representations. The table in the example introduced in the main text acquires value based on the context and the goal of the decision-maker. In the construction of abstract, goal-dependent value representations, there are three main steps. First and foremost, the brain (or the artificial system doing the valuation) needs to build an abstract representation from purely sensorial information. For instance, if the goal is to evaluate the usefulness of a table for a meeting, it might focus on a subset of features such as the size, the design, the colour, etc. The selection of these features for abstraction is operated (mainly, but not exclusively by) attention and memory. Finally, if the goal changes, for instance, because of a sudden earthquake, the brain needs to update this value representation, extracting new/different features such as sturdiness, height, and location, while disregarding other features irrelevant to the new behavioural goal (i.e., colour, design).

Box 1. How artificial intelligence (AI) agents learn from multiple goals

Reinforcement learning (RL) is generally seen as a 'single objective' algorithm [108]. The agent seeks to find the best policy to solve one problem (e.g., exit a maze) by maximising future returns or minimising future punishments (integrating a single scalar on each outcome). While humans and other animals undoubtedly use this type of RL [109–111], practical problems in real life are often complex and contain multiple rewards or objectives in parallel (e.g., a drug needs to be effective, have as few side effects, be inexpensive, easy to produce….).

These intuitions have led to important algorithmic developments in AI. In the 1990s, hierarchical RL expressed as mixture-of-experts was developed to deconstruct complex problems through a divide and conquer strategy. There, a controller allocates a new case to one or a few subnetworks (experts). This architecture was first used for pattern recognition problems, such as phoneme discrimination [112] and vowel recognition [113] and, later, for applications in robotics and motor control [114,115]. Hierarchical or modular RL and multi-agent RL are similar algorithms designed to solve tasks made of multiple subproblems.

In contrast, multiobjective (or multigoal) reinforcement learning (MORL) [116,117] confronts classes of problems where the agent explicitly trades multiple (often conflicting) objectives. The single scalar reward limitation arising in standard RL is overcome by defining vectors of reward signals, one for each objective. Values are encoded separately for each objective. The agent learns the relative importance of fulfilling each objective using preference weights. These weights can be dynamic [118], increasing the flexibility of the agent to changing or *a priori* unknown conditions. MORL aims to maximise rewards over all objectives simultaneously, weighted by their relative importance. Thus, in MORL there exists a set of optimal policies, called the Pareto Front (from the Italian economist Vilfredo Pareto), outside of which no other policy is equal or better in all objectives [117]. However, MORL is computationally demanding and introduces new problems, such as the horizon of objectives (the number of objectives can quickly grow to become intractable) and the nontrivial solutions to Pareto optimality.

The development of new multigoal algorithms might benefit from a deeper understanding of how humans solve multigoal problems. It has been documented that in multitask RL, people evaluate a set of previously learned policies considering the currently valid reward function and task states [119]. This finding is particularly interesting because it suggests a single abstract value representation can be paired with multiple policies. Others have found that people are more flexible in updating goal-directed task feature processing toward rewards than toward punishments in dynamic multigoal scenarios [120].

should be part of your state representation [10] (for an extensive discussion, see [11]). In machine learning, this problem can be often circumvented by hand-crafting the correct state or directly using low-level inputs as a state, although this latter approach requires a huge amount of training data [3].

In the next sections, we will discuss some mechanisms that the brain might implement to construct these subjective value representations by abstracting the correct state to fulfill a goal. Most of these mechanisms can apply equally to: (i) learning problems, and (ii) simple everyday goal-value decisions like the one presented in the opening example.

Abstractions

We define abstractions as high-level compact representations that are transferable to new situations, allowing agents to quickly adapt when goals or the environment change. We will focus on how abstractions are formed to fulfill a precise goal and guide value-based choice.

Reducing dimensions

A key step in building useful abstract value representations is **dimensionality reduction**, which has been studied in fields ranging from computational linguistics to visual processing [12,13], statistical learning [14,15], categorizations [16–18], and concepts or **rules** [19,20]. For example, humans can recognise items that belong to the same category (e.g., chairs, food) even when the visual inputs are extremely different by focusing only on those dimensions that are important for the categorization. This powerful cognitive tool may be a critical ingredient for the brain to overcome, during learning, the so-called **curse of dimensionality** [21] by reducing the complexity of the underlying representations [22].

We recently investigated how humans learn to solve decision problems based on abstractions [23]. In the task, hidden rules defined what information was relevant or irrelevant. Rules could be learnt via two different strategies: (i) using all (redundant) stimuli features; or (ii) using abstraction (integration of relevant information alone). People who based their decision strategies on abstractions learnt faster and were more confident about their performance. Simulations in RL artificial agents replicated accelerated learning. This ability was underpinned by value signals in the ventromedial prefrontal cortex (vmPFC), an area prioritised during abstraction. Abstractions were learnt: at the beginning people largely used raw task features, while later they relied on abstract reasoning. Others have made similar findings [24], for example, during navigation the brain also learns the relevant abstractions (e.g., a stylized representation of a maze [25,26]).

The field of **cognitive control** has made the most progress in understanding how a behavioural demand or goal triggers a reduction in dimensions. However, most experimental tasks used in this field (such as the Stroop task, Eriksen flanker task, and n-back task [27]) are relatively simple compared with the tasks used in value-based decisions. This is because experiments in cognitive control are mostly concerned with understanding the input–output remapping in response to task demands (i.e., goal manipulation). In contrast, the field of value-based choice adopts a different approach, in which tasks tend to be characterised by rich stimuli and complex structures. Moreover, in the majority of value-based tasks, the goal is generally singular and straightforward (to maximise a numeric reward) and is seldom manipulated during the task. A direct comparison of these two approaches can be found in an in-depth review by Frömer & Shenhav [28].

Are low-dimensional abstract representations always useful? Theoretical and empirical work has demonstrated a trade-off between low and high dimensionality of representations [29–31]. While low-dimensional representations are robust to noise and changes in input, they fail to separate similar inputs. This cost is probably inconsequential in value tasks, which, unlike perceptual discrimination, use stimuli that are perceptually easy to separate (e.g., a banana and ice cream). But given that stimuli are composed of many features/dimensions, a more serious concern in value-based choice is the possibility of compressing the wrong dimensions. If the abstraction is built at the wrong level, such as overly simple features, then the process becomes slow and inefficient [23,32]. We will come back to this problem later and discuss putative mechanisms that can monitor the reliability of the current abstraction given the agent's overarching goal(s).

Goals also define the hierarchy of abstraction and along which dimensions abstraction should develop. For example, abstractions can be: (i) at the stimulus level, a representation of colour and texture, ignoring shape and size; or (ii) at the task or conceptual level, such as multiple stimuli linked by a rule, or category; and even (iii) in time, merging information across trials to represent task structure. Value-based abstractions are thus forged through the integration of sensory stimuli with motor and/or memory representations. How much do we know about how this integration is instantiated at the neural and algorithmic levels?

Mixed selectivity

In many brain areas, most notably in the prefrontal cortex (PFC), neurons often display complex firing responses, apparently mixing information sources such as cues, stimuli, **contexts**, rewards, etc., a phenomenon named **mixed selectivity** (Figure 2A). Theoretical work has shown that populations of mixed-selective neurons can control the trade-off between generalisation and discrimination, mapping onto high and low level of abstraction [33]. High-dimensional neural representations with mixed selectivity allow a simple linear readout to generate a vast number of unique potential responses. In contrast, neural representations based solely on specialised neurons (high selectivity) will lead to low-dimensional neural

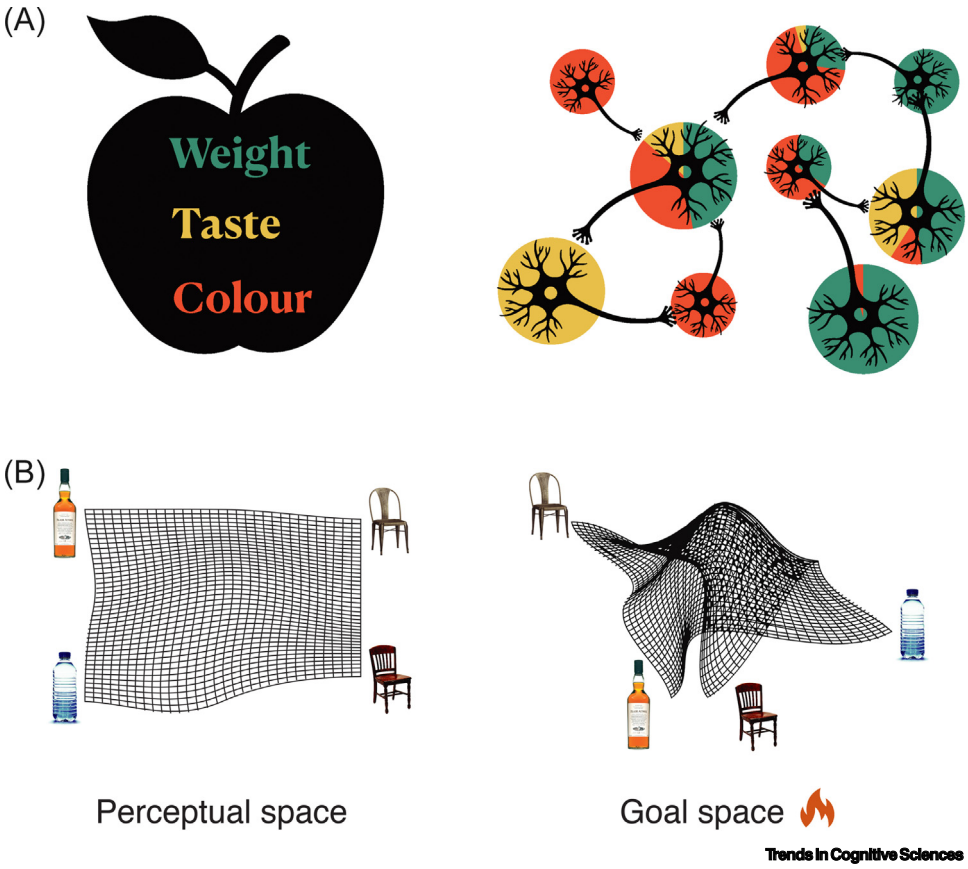

Figure 2. Mixed selectivity and goal-directed abstraction in decisions. Mixed selectivity is a possible algorithmic implementation that allows an optimal mixture of low/high-dimensional coding and representations. (A) Graphical illustration of neurons with various levels of selectivity, from pure selective to a single task variable (e.g., yellow neuron exclusively coding 'taste') to mixed selectivity over multiple variables (e.g., red-green neuron coding a mixture of 'colour' and 'weight'). (B) Participants were asked to rate the usefulness of common objects under two different goals. In the first goal, participants had to judge how useful an item was to light a fire, and in the second goal, participants had to do so with respect to how well an item could be used to anchor a raft. In this experiment, perceptually similar stimuli (such as a wooden chair and a metal chair) flexibly rearrange into new goal-dependent representations that emphasise goal-relevant features. If one's goal is to choose a burnable object, a wooden chair and bottle of whisky will be closer in the representational space than a metal chair and a (perceptually more similar) wooden chair. Both panels adapted from [42].

patterns [29]. In humans, high-dimensional neural representations predict effective learning [34] and better episodic memory [35].

So far, we have discussed abstraction as a process that reduces the dimensionality of information. However, this idea seems to contradict the experimental and theoretical evidence introduced in the previous paragraph, exemplified by neural recordings showing these representations are simultaneously abstract and high-dimensional [36]. How can this apparently counterintuitive mechanism be advantageous? One reason is that it leads to an agent that avoids large information loss because it does not 'relearn' dimensions and instead can resurface information as needed for on-the-fly computations. Therefore, the representations used to guide behaviour are functionally low-dimensional, but the underpinning neural code might be high-dimensional (for further reading, see [29,31,37]).

While PFC neurons show complex and mixed response profiles, each cell's activity nevertheless primarily covaries with a single task parameter, including value, as extracted from computational

models of choice behaviour [38]. Thus, prefrontal neurons may implement a multiplexed coding principle, the readout of which depends on the demands of downstream neurons or circuits. Importantly, neurons also have other means of controlling the dimensionality of representations (besides mixed selectivity), such as synchronisation across neurons [39] and oscillatory dynamics [40]. It remains to be further investigated how these different mechanisms are connected and their role in value-based abstractions.

On-the-fly abstractions for value-based choice

Beyond learning, in most value-based choices abstractions can be summoned or modified on-the-fly [41]. Changes in goals quickly rearrange the neural representations, neural manifolds, and the compression of relevant dimensions. A recent study from one of our labs has shown that such rearrangements are virtually instantaneous and unfold as an automatic process [42]. In this work, participants were asked to picture themselves in an emergency and imagine using everyday items to fulfill two goals: lighting a fire or anchoring a boat. Using representational similarity analysis, the study showed that activity patterns in visual regions coding for object representations were clustered according to their perceptual features. Meanwhile, in regions of the PFC (usually associated with value computation), the pattern of representation similarity was reshaped by the goal. For example, if the goal was burning, the representation of a wooden chair was more like a bottle of whiskey (both flammable items) than a metal chair (Figure 2B). In Box 2, we discuss the puzzling existence of aesthetic values as an extreme form of value abstractions that is portable across wildly different scenarios.

Information selection

As hinted in the previous sections, a critical step in building abstract (compact) value representations is selecting the correct information. Here, we focus on the role of attention and memory.

Box 2. Aesthetic values

How can we explain aesthetic values in the framework we proposed? It is difficult to imagine how listening to the Chaconne from Bach or watching the northern lights is a useful pursuit (Figure I). What is even more puzzling is that aesthetic values activate the same brain network [121] that is involved in the goal-dependent value representations discussed throughout this article. How aesthetic preferences develop has flummoxed philosophers for centuries [122,123]. Recently, neuroscientists have started to tackle this problem experimentally [124,125]. An ingenious recent study used deep convolutional networks to show how low-level perceptual features are combined and integrated to predict subjective preference for visual art [126]. Nevertheless, the relation between aesthetic and goal-dependent value is still opaque. We briefly present some ideas that admittedly are very speculative, but we hope they might inspire future computational and experimental work.

In everyday language, we apply aesthetic categories such as ‘beautiful’ or ‘ugly’ to describe different phenomena across many different domains. A chess player might say ‘this position is beautiful’; a mathematician ‘this proof is correct but ugly’. What does a beautiful chess position have in common with an elegant mathematical proof? Is it just imprecise use of language or is there something more to it? We suggest that aesthetic judgement is a form of (very) high-level abstraction that captures some statistical regularity in the environment (both natural but also cultural). Because of the high degree of abstraction, these representations might not be fully accessible to awareness or verbal reports beyond the very coarse definitions that we use in aesthetic judgement (beautiful, ugly, graceful…). These abstract aesthetic value representations can capture commonalities and relations that exist in the cultural and natural environment, while at the same time discarding lower-level features specific to a given situation or modality. These representations are therefore extremely portable, allowing for an impressive degree of generalisation and reducing the amount of training required by the human brain. For example, if one learns complex relations in the visual domain (e.g., a painting), one might be able to learn more easily similar relations in the musical domain (e.g., music). These aesthetic values when formed are not static, but they get updated with experience. If one is used only to classical music from the Baroque or romantic period, one will probably find the Rite of Spring of Igor Stravinsky ugly due to the strident dissonances. However, with time, we intuitively discover new hidden structures and relations. While we are unable to describe this verbally, these newly discovered relations might change our aesthetic judgement about that piece of music. The understanding and appreciation of these new abstract relations can in turn affect the aesthetic valuation of a painting by Rothko or an unusual dessert by Jordi Roca.

Figure I. Aesthetic values. (Top) Johann Sebastian Bach score of the incipit of the Partita No. 2 for solo violin (BWV1004). (Bottom) A photo of the northern lights (aurora borealis).

Attention

Attention is arguably the main mechanism the brain uses to prioritise information processing. During perception, (top-down) attention modifies sensory representations. This happens in vision, where attention strengthens the representation of attended stimuli, for example, by increasing the selectivity of population responses [43] or enhancing local fMRI signals [44–46]. Attention effects on sensory representations require a direct prefrontal control of sensory neurons [47,48], including PFC-driven dampening of visual distractors [49].

Something analogous might be triggered by goals. In the study presented in Figure 2B, the usefulness of an item changed its internal representations, even in occipital areas [42], which were incrementally recruited according to the item's usefulness. This finding is reminiscent of the value-driven attentional capture of high-value sensory features [50]. More broadly, neural coding in the sensory cortex is sensitive to the effect of reward [51–53]. Orbitofrontal cortex neurons have been shown to exert direct control over sensory cortices in value-based choices [54,55].

These results across sensory domains have been interpreted in different ways. Any good mechanism needs to explain a two-way interaction between sensory cortices and PFC: (i) a top-down strengthening of relevant sensory representations; (ii) prioritisation of the 'relevant' sensory features that are combined to form an abstraction in PFC on which learning or choice operates.

We recently studied how reward guides attention selection and how it relates to the formation of abstract representations during learning [23]. We tested the hypothesis that, during learning, reward should operate a similar attention-capture mechanism on sensory features relevant for abstraction. Based on fMRI data analysis, we found the area of the vmPFC encoding value strength (as calculated by an RL model) was functionally coupled with the occipital and lingual gyri in episodes that lead to reward. The strength of the coupling was relevant to behaviour, as it was correlated with participants' ability to build abstractions and learn more efficiently. To investigate the directionality of this coupling, we exogenously paired the sensory neural representation of a target feature with monetary reward using neurofeedback. In this procedure, each participant received a reward every time we detected a neural multivariate activity pattern classifying a specific visual feature. This was done in the absence of visual stimuli and with participants unaware of which feature was paired with reward. In a follow-up test, we showed that participants learnt faster when the feature tagged with a reward (via neurofeedback) was part of the correct abstraction.

But the relationship between attention and value has been intensely studied over the past decade, especially in the context of simple choice (i.e., outside learning). A well-known empirical finding is that people tend to look longer at more valuable items. In addition, the time spent attending to an item is proportional to the probability that the item will be eventually chosen. The intuition was that, during value comparison, attention boosts the value of the attended item either by amplifying its magnitude [56–58] or by shifting its baseline upwards by a constant amount [59]. However, often (with a few exceptions [60,61]) the goal of the task was not directly manipulated. Therefore, choosing which snack to eat (i.e., goal value) was equivalent to choosing the snack I like most (i.e., hedonic value). In a recent study, we used a simple goal-framing manipulation in which participants were sometimes asked to choose the least preferred item. We showed that, contrary to the commonly held view, attention does not boost reward processing *per se* but selectively prioritises information relevant to achieving the current goal [62]. This effect also generalised to perceptual tasks. The implication of these findings is that behavioural goals shape the type of information that is processed and integrated from the very incipit of the evaluation process. An analogous mechanism has been detected in recent studies on confirmation bias [63–65]. The simple (and often inconsequential) commitment to a choice influences how information is thereafter acquired and processed.

Memory

While attention is crucial for information selection and prioritisation, memory is arguably the primary source of information from which abstract value representations are constructed. For example, in the studies presented in the previous paragraph, it is questionable that spending more time looking at a picture of a well-known snack item would provide more sensory information to compute the value of the item. It is more likely that information is retrieved from memory and (for unknown reasons) eye gaze provides a window into the internal sampling process [66,67]. This makes the problem of dimensionality reduction that we introduced earlier in the context of perception even more severe. There are effectively endless memories stored in the brain and only a tiny fraction of them are important for the current decision. An important feature of memories is the way they can be structured in schemas, which are compositions of primary

information [68], allowing faster access to target information [69]. For example, it is easier to recall the name of a work colleague in an office environment than in a bar.

We suggest that there is a fundamental similarity at the computational level between schema and the abstract value representations so far discussed. This would explain why lesions in the same brain region (vmPFC) impair both the formation and retrieval of schematic memory and value-based choice [70,71]. In support of this view, recent studies have shown that vmPFC plays a critical role that requires a novel combination of information in an abstract representation to compute value [41,72].

The exact relation between schemas and value is not yet fully understood. However, one can speculate that abstract value representations are constructed similar to schemas. The goal provides the context to retrieve information, aiming to fulfill a specific behavioural goal. Akin to perception, the goal constrains the retrieval content; retrieval in turn could initiate a circuit reverberation, amplifying only what is relevant (Figure 3). Like the process of filtering and prioritising perceptual information, attention also modulates the memory retrieval process [73]. A new fruitful research agenda has started to dissect the interaction between attention and memory [74–76]; this will shed light on the algorithmic and neural mechanisms involved in the formation of goal-dependent value representations.

Moreover, memory is central in RL algorithms. A promising approach, known as 'successor representation RL' (SR), stores long-term predictions of future states that will be visited. SR shares the simplicity of model-free RL but is still capable of approximating (some of) the flexibility

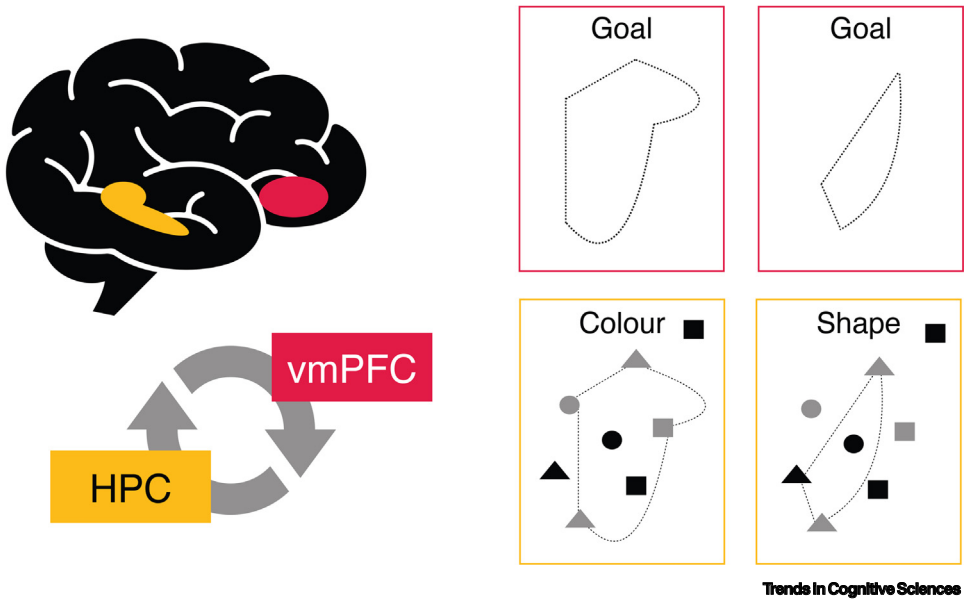

Figure 3. The link between ventromedial prefrontal cortex (vmPFC) and hippocampus (HPC) in evaluating and updating goal-dependent values. The vmPFC encodes high-level schemas or maps that are described here as abstract goal-value mappings. These maps guide the HPC in the retrieval (or encoding) of specific relevant elements (in the example, grey colour objects vs. triangles), ignoring the irrelevant memories. This can trigger a recursive process in which the HPC amplifies only what is relevant and in doing so shapes the goal-dependent map built by vmPFC.

of model-based RL [77,78]. An open question is whether goals modulate the encoding and/or the retrieval of these future states represented by the SR algorithm.

Control of the representations: monitoring and updating

Goal-dependent values must adapt when situations change and an abstraction fails to be useful. In the table example, one might notice that the wood is starting to crack under the weight of the debris. This observation makes the table less valuable (useful) for one's protection and one might seek refuge in a door frame instead. The brain needs a two-step mechanism that monitors the quality of a value abstraction and updates/replaces the representation when it ceases to be useful.

The human brain has developed a sophisticated mechanism for monitoring internal representations and adjusting behaviour. The ability to evaluate our own thoughts or performance (i.e., metacognition) is often measured in the lab using confidence reports. In computational terms, metacognition can be quantified as the mutual information between the accuracy of the agent's choices and their confidence reports [79]. Recent work from ours and other labs has shown that, through metacognition, people can also monitor their value representation and adjust behaviour [80–85]. At the neural level, confidence signals can be tracked across different subregions of the prefrontal and cingulate cortices, including vmPFC [86]. Overlapping signals for value and confidence are associated with control of behavioural measures such as choice and reaction time. These findings point to a deeper algorithmic role of confidence in adaptive behaviour and the control of internal representations. In line with this idea, reliability (measured experimentally using confidence reports) appears to be an intrinsic feature of value [83,85,87,88]. Usefulness,

Box 3. Algorithms for monitoring, controlling, and updating abstract value representations

Hierarchical switch-evidence variable

In an influential study conducted by Sarafyazd and Jazayeri on monkeys performing a task that required credit assignment of an error source, it was suggested that confidence controls the accumulation of evidence in favour of a behaviour switch to the alternative strategy [127]. This was achieved using a simple but compelling hierarchical algorithm that uses a 'switch evidence variable' as input to a step function, in that it signals 0 (stay) if the switch evidence is below a threshold and 1 (switch) otherwise (Figure I, top-left). The critical question is how the switch variable is computed and used by the brain. Experimental findings indicated the dorsomedial frontal cortex accumulated switch evidence, while the dorsal anterior cingulate cortex operated downstream to determine the behaviour course [127].

Concurrent behavioural strategies selection

A related line of research on reasoning in humans has put forward an influential cognitive model postulating that the brain performs two parallel computations to solve the exploit/explore problem, which can be adapted to our case as: 'exploit' = continue using the current abstraction; 'explore' = change abstraction [128,129]. This model considers that humans evaluate a few alternatives and use hypothesis testing to select a strategy (Figure I, bottom-left). This resembles a decision-theoretical idea in economics, in which agents engage with one hypothesis at a time (i.e., an abstract representation in our case). When this hypothesis stops being appropriate, it is rejected and replaced by the next best one that has been selected, following Bayes' rule [130].

Meta-learning

Similar control mechanisms must apply when the representations are learnt. Meta-learning is a suitable candidate mechanism in RL to perform a similar type of control. In meta-learning the agent learns hierarchical aspects of a task, over different timescales [108,131,132] (Figure I, top-right). The **cortico-thalamic loops** linking the PFC and basal ganglia/thalamus provide the neural underpinning for meta-learning [133,134]. Artificial neural networks designed to mimic this dual neural component model display efficient episodic learning and hallmarks of classic RL observations [133]. This style of learning allows one to flexibly map different levels of representations, finding the most appropriate abstraction to fulfil the current goal.

Mixture-of-experts

The fields of AI and robotics have developed other, related models, such as the mixture-of-experts architecture [113,135]. Although this model fell out of favour due to over-parameterized and excessively constrained implementations, new work is showing its merit as a neurocognitive mechanism [23,26,136,137]. Mixture-of-experts can flexibly integrate multiple representations, including multiple levels of abstractions at once: an internal (implicit) controller selects the best expert(s) based on a 'responsibility' signal (derived from priors and prediction errors from each expert) [23,114,115] (Figure I, bottom-right). Note the strong similarity between responsibility and reliability discussed earlier (concurrent behavioural strategies selection). We suggest abstract value representations serve a crucial function: selecting the right expert(s) (i.e., the right task set or representation).

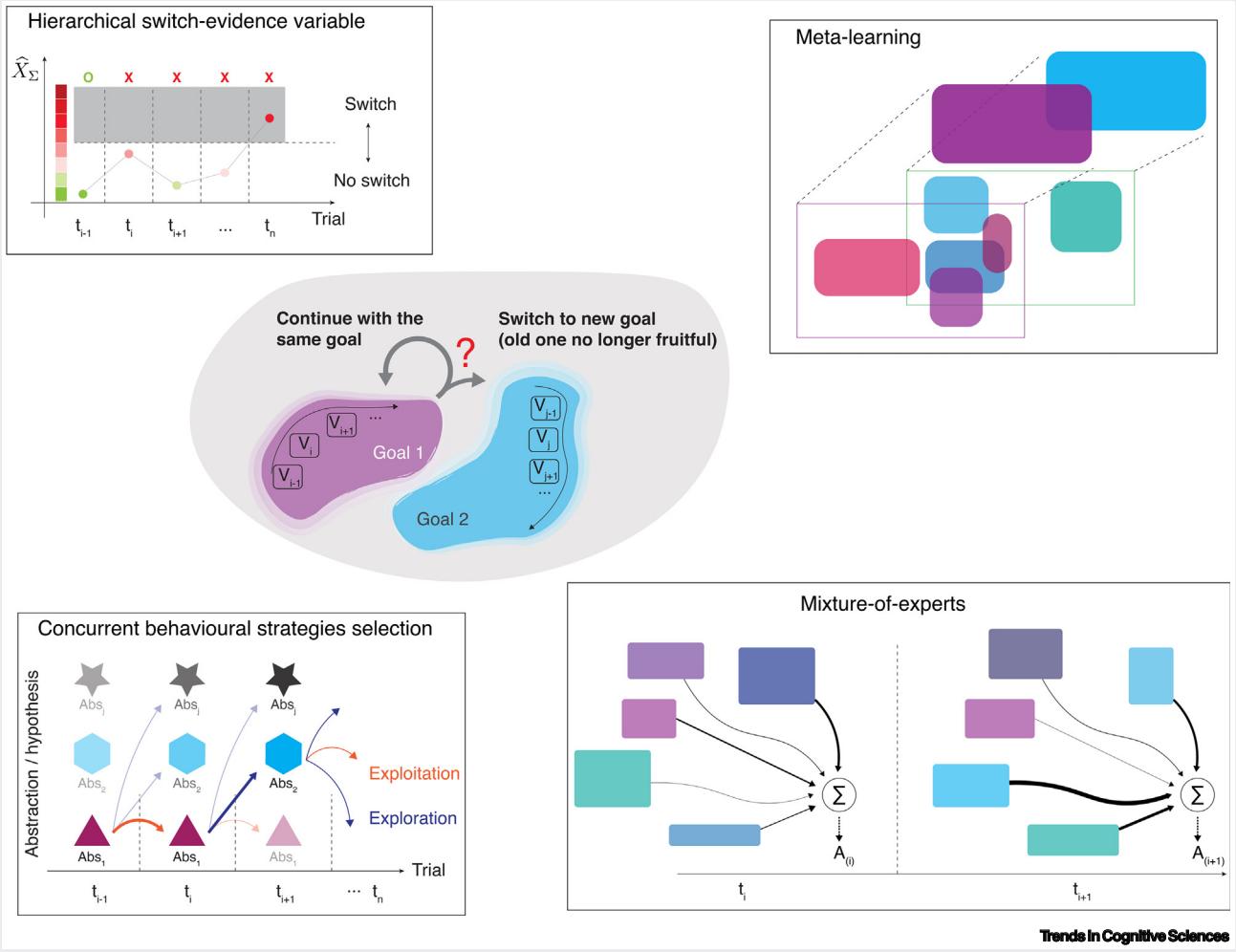

Figure I. Algorithms for monitoring and control/update of abstract value representations. Values are computed over time according to a first goal (goal 1). Changes in the environment, context, rules, etc. should lead the decision-maker to re-evaluate the goal and the value construction process, to best fit current demands. This arbitration problem is critical and defines abstract goal value. We describe four possible algorithms, which were originally developed in other contexts. Top-left: hierarchical switch-evidence variable [23]. This model updates the switch evidence X_Σ based on the history of previous outcomes (reward vs. error), and the agent's belief about the stimulus and rule. After rewarded trials (green circle), X_Σ is reset to zero; after each error (red cross), X_Σ increases. When X_Σ breaches a threshold (broken line), the model switches the rule. Bottom-left: concurrent behavioural strategies selection. Shapes represent abstractions. Reliabilities of monitored representations are inferred from action outcomes and used to switch into exploration (test alternative hypotheses) when abstraction(s) becomes unreliable. Exploration periods end when a new abstraction becomes reliable. Top-right: meta-learning. A set of abstract value representations are learned across goals, over different timescales, and hierarchically structured. Bottom-right: mixture-of-experts architecture. Each expert (coloured rectangles) represents an abstraction. Switching between abstractions/representations is done by reweighting the selective importance of each expert in determining the agent's actions.

reliability, or value might be partially overlapping psychological constructs that characterise various aspects of the same algorithm dedicated to implementing actionable goals through the generation, monitoring, and switching of structured abstract representations [42,89].

However, monitoring the quality of representations is just the first step. The brain also needs a mechanism that switches and selects new representations when needed. Our understanding of how the brain achieves this stems from studies on rules, context, and executive functions in memory or cognitive control. We suggest that the same principle could apply to goal-dependent

abstract value representations. How might this process unfold at the neurocomputational level? An influential study revealed that context-dependent neural computations in the PFC generated separable (orthogonal) representations through recurrent dynamics, as the product of a single process of information selection and integration [90]. Considering the multiplexed nature of single PFC neurons' activity, coupled with line attractor and selection vector at the population level, the function of the PFC may be to generate, link, and select separate representations. The importance of the PFC in controlling representations was underscored by another study that found PFC operated as a domain-general controller in working memory, both selection of memory dimension and attention to stimuli [91]. Other areas, such as the visual or parietal cortex, operated each process independently.

Studies on rules, and especially rule search, also provide insight into how the brain controls inflow, processing, and abstraction to separate task-relevant from irrelevant information. Activity in the medial PFC predicted participants' future strategy changes [92]. More specifically, when this region started encoding information irrelevant to the main explicit rule, but necessary for an undisclosed rule, participants were more likely to discover and switch to the new and more efficient rule.

Similarly, other regions of the frontal cortex, such as frontopolar cortex, can play a key role in monitoring the quality of a choice during metacognitive judgements in both perceptual [93] and value domains [83,94]. We believe that it is not an accident that this same region has been shown in controlling arbitration between different learning strategies [95], managing competing goals [96], and counterfactual judgements [93]. Although a full picture of the underpinning neural computation that controls and updates abstract value representations remains elusive, some candidate algorithmic implementations are discussed in Box 3.

Concluding remarks

These are exciting times in which rapid developments in AI algorithms are fostering the integration of different domains of psychology and cognitive neuroscience. Such integration will undoubtedly help to uncover general computational principles, which apply both to the human brain and AI. However, different fields of cognitive neuroscience have often interpreted their results through narrow lenses, neglecting data and interpretations from other fields. A paradigmatic example is the vmPFC, whose function has been linked to the computation of economic value, reward, and confidence in the field of neuroeconomics [97,98], schemas in the field of memory [68,99], regulation of impulsivity in the field of cognitive control [100], prioritisation of information in working memory [101], and more recently to the formation of cognitive maps [102,103]. We believe that it is possible to reconcile these (superficially) quite different findings under a common computational language. Other scholars have been advocating a similar research program [28,102,104].

In this review, we have suggested that value is a functional construct that does not necessarily overlap with reward. These differences are often blurred in experimental designs in which the goal of the participant is to maximise simple numerical rewards. As we discussed, recent work has started to disentangle these concepts by manipulating the goal of the task. However, it is not always clear what constitutes a context, a rule, or a goal manipulation. More experimental and computational work is required to tease apart these overlapping psychological constructs.

Finally, we still lack a satisfactory neurocomputational account of how an agent endogenously sets and changes its own goals (see Outstanding questions). Most learning algorithms implicitly assume that agents have one goal and receive exogenous rewards (often scalar numbers) from the environment. This scenario is unrealistic for real agents operating in real environments.

Outstanding questions

How does hedonic reward interact with goal value? We are still missing a clear understanding of the computational origin and the objective of hedonic value and its interaction with functional abstract values (i.e., usefulness) that we presented in this review.

At the neuromodulatory level, how do prediction errors mediated by the dopaminergic system interact with the opioid system associated with the hedonic state of an agent?

How do humans set their own goals? What is the algorithm humans use to arbitrate amongst different competing goals? While in most studies goals (usually one at a time) are set by the experimenter, this is often not the case for a real decision-maker operating in a real environment. We still do not have a clear algorithmic answer on how human decision-makers choose among competing goals which one to pursue and when to switch to a different goal.

What are the differences between goal, context, or rule? Do these terms map onto different psychological constructs? Do goals require a sense of agency while rules are exogenous to the decision-maker? If this is the case, we will then need to understand how to capture these differences algorithmically and how they are implemented neurally.

Does the brain map externally imposed contexts or rules as an internal goal, such that the same mechanism can be used across scenarios?

We have only just begun to consider how reward functions might be designed in biological organisms [105]. The notion of homeostatic RL [106,107], where the value of an item depends on the physiological needs and homeostatic balance of the agent, is a timely new research direction. By combining ideas drawn from different fields, sophisticated experimental designs, and the formalism of computational models, we are reaching a better understanding of the stupefying human ability to flexibly respond to everchanging behavioural demands.

Acknowledgments

B.D.M. is supported by the Japan Trust International Research Cooperation Program of National Institute of Information and Communication (NICT), and by a Google Faculty Research Award. A.C. is supported by the Ikegaya Brain-AI Hybrid ERATO program (grant number JPMJER1801) from the Japan Science and Technology Agency (JST); by the Innovative Science and Technology Initiative for Security (grant number JPJ004596) from ATLA, Japan; by JSPS KAKENHI (grant number JP22H05156), Japan. We would also like to thank Alice De Martino for her help in proofreading the manuscript.

Declaration of interests

No interests are declared.

References

1. Duncan Luce, R. and Raiffa, H. (1989) Games and Decisions: Introduction and Critical Survey. Courier Corporation
2. Sorokin, I. *et al.* (2015) Deep attention recurrent Q-network. *arXiv* Published online December 5, 2015. https://doi.org/10.48550/arXiv.1512.01693
3. Mnih, V. *et al.* (2015) Human-level control through deep reinforcement learning. *Nature* 518, 529–533
4. Silver, D. *et al.* (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489
5. Whittington, J.C.R. *et al.* (2020) The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell* 183, 1249–1263
6. Behrens, T.E.J. *et al.* (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100, 490–509
7. Mattar, M.G. and Lengyel, M. (2022) Planning in the brain. *Neuron* 110, 914–934
8. Witkowski, P.P. *et al.* (2022) Neural mechanisms of credit assignment for inferred relationships in a structured world. *Neuron* 110, 2680–2690
9. Cortese, A. *et al.* (2019) The neural and cognitive architecture for learning from a small sample. *Curr. Opin. Neurobiol.* 55, 133–141
10. Ghetti, S. and Coughlin, C. (2018) Stuck in the present? Constraints on children's episodic prospection. *Trends Cogn. Sci.* 22, 846–850
11. Niv, Y. (2019) Learning task-state representations. *Nat. Neurosci.* 22, 1544–1553
12. Poggio, T. and Bizzi, E. (2004) Generalization in vision and motor control. *Nature* 431, 768–774
13. Poggio, T. *et al.* (2004) General conditions for predictivity in learning theory. *Nature* 428, 419–422
14. Turk-Browne, N.B. *et al.* (2009) Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* 21, 1934–1945
15. Schapiro, A.C. *et al.* (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* 22, 1622–1627
16. Pan, X. *et al.* (2008) Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nat. Neurosci.* 11, 703–712
17. Pan, X. and Sakagami, M. (2012) Category representation and generalization in the prefrontal cortex. *Eur. J. Neurosci.* 35, 1083–1091
18. Freedman, D.J. *et al.* (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316
19. Saez, A. *et al.* (2015) Abstract context representations in primate amygdala and prefrontal cortex. *Neuron* 87, 869–881
20. Wallis, J. *et al.* (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956
21. Bellman, R. (1957) Dynamic Programming. Princeton University Press
22. Ponsen, M. *et al.* (2010) Abstraction and generalization in reinforcement learning: a summary and framework. In *Adaptive Agents and Multi-Agent Systems IV* (Taylor, M.E. and Tuyls, K., eds), Springer-Verlag
23. Cortese, A. *et al.* (2021) Value signals guide abstraction during learning. *eLife* 10, e68943
24. Schuck, N.W. and Niv, Y. (2019) Sequential replay of nonspatial task states in the human hippocampus. *Science* 364, eaaw5181
25. Miller, A.M.P. *et al.* (2019) Retrosplenial cortical representations of space and future goal locations develop with learning. *Curr. Biol.* 29, 2083–2090
26. Ho, M.K. *et al.* (2022) People construct simplified mental representations to plan. *Nature* 606, 129–136
27. Gratton, G. *et al.* (2018) Dynamics of cognitive control: theoretical bases, paradigms, and a view for the future. *Psychophysiology* 55, 3
28. Frömer, R. and Shenhav, A. (2022) Filling the gaps: cognitive control as a critical lens for understanding mechanisms of value-based decision-making. *Neurosci. Biobehav. Rev.* 134, 104483
29. Fusi, S. *et al.* (2016) Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74
30. Rigotti, M. *et al.* (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590
31. Badre, D. *et al.* (2021) The dimensionality of neural representations for control. *Curr. Opin. Behav. Sci.* 38, 20–28
32. Eckstein, M.K. and Collins, A.G.E. (2020) Computational evidence for hierarchically structured reinforcement learning in humans. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29381–29389
33. Barak, O. *et al.* (2013) The sparseness of mixed selectivity neurons controls the generalization-discrimination trade-off. *J. Neurosci.* 33, 3844–3856
34. Tang, E. *et al.* (2019) Effective learning is accompanied by high-dimensional and efficient representations of neural activity. *Nat. Neurosci.* 22, 1000–1009
35. Sheng, J. *et al.* (2022) Higher-dimensional neural representations predict better episodic memory. *Sci. Adv.* 8, eabm3829
36. Bernardi, S. *et al.* (2020) The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* 183, 954–967
37. Vaidya, A.R. and Badre, D. (2022) Abstract task representations for inference and control. *Trends Cogn. Sci.* 26, 484–498
38. Hirokawa, J. *et al.* (2019) Frontal cortex neuron types categorically encode single decision variables. *Nature* 576, 446–451
39. Hoang, H. *et al.* (2020) Electrical coupling controls dimensionality and chaotic firing of inferior olive neurons. *PLoS Comput. Biol.* 16, e1008075

40. Wutz, A. *et al.* (2018) Different levels of category abstraction by different dynamics in different prefrontal areas. *Neuron* 97, 716–726
41. Barron, H.C. *et al.* (2013) Online evaluation of novel choices by simultaneous representation of multiple memories. *Nat. Neurosci.* 16, 1492–1498
42. Castegnetti, G. *et al.* (2021) How usefulness shapes neural representations during goal-directed behavior. *Sci. Adv.* 7, eabd5363
43. Martinez-Trujillo, J.C. and Treue, S. (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol.* 14, 744–751
44. Somers, D.C. *et al.* (1999) Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1663–1668
45. Watanabe, M. *et al.* (2011) Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science* 334, 829–831
46. Guggenmos, M. *et al.* (2015) Spatial attention enhances object coding in local and distributed representations of the lateral occipital complex. *Neuroimage* 116, 149–157
47. Barceló, F. *et al.* (2000) Prefrontal modulation of visual processing in humans. *Nat. Neurosci.* 3, 399–403
48. Noudoost, B. and Moore, T. (2011) Control of visual cortical signals by prefrontal dopamine. *Nature* 474, 372–375
49. Cosman, J.D. *et al.* (2018) Prefrontal control of visual distraction. *Curr. Biol.* 28, 414–420
50. Anderson, B.A. *et al.* (2011) Value-driven attentional capture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10367–10371
51. Arsenault, J.T. *et al.* (2013) Dopaminergic reward signals selectively decrease fMRI activity in primate visual cortex. *Neuron* 77, 1174–1186
52. Henschke, J.U. *et al.* (2020) Reward association enhances stimulus-specific representations in primary visual cortex. *Curr. Biol.* 30, 1866–1880
53. Watanabe, M. (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382, 629–632
54. Banerjee, A. *et al.* (2020) Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* 585, 245–250
55. Liu, D. *et al.* (2020) Orbitofrontal control of visual cortex gain promotes visual associative learning. *Nat. Commun.* 11, 2784
56. Krajbich, I. *et al.* (2010) Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* 13, 1292–1298
57. Krajbich, I. and Rangel, A. (2011) Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13852–13857
58. Smith, S.M. and Krajbich, I. (2019) Gaze amplifies value in decision making. *Psychol. Sci.* 30, 116–128
59. Cavanagh, J.F. *et al.* (2014) Eye tracking and pupillometry are indicators of dissociable latent decision processes. *J. Exp. Psychol. Gen.* 143, 1476–1488
60. Frömer, R. *et al.* (2019) Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. *Nat. Commun.* 10, 4926
61. Kovach, C.K. *et al.* (2014) Two systems drive attention to rewards. *Front. Psychol.* 5, 46
62. Sepulveda, P. *et al.* (2020) Visual attention modulates the integration of goal-relevant evidence and not value. *eLife* 9, e60705
63. Talluri, B.C. *et al.* (2018) Confirmation bias through selective overweighting of choice-consistent evidence. *Curr. Biol.* 28, 3128–3135
64. Kaanders, P. *et al.* (2022) Humans actively sample evidence to support prior beliefs. *eLife* 11, e71768
65. Palminteri, S. and Lebreton, M. (2022) The computational roots of positivity and confirmation biases in reinforcement learning. *Trends Cogn. Sci.* 26, 607–621
66. Shadlen, M.N. and Shohamy, D. (2016) Decision making and sequential sampling from memory. *Neuron* 90, 927–939
67. Shushruth, S. *et al.* (2022) Sequential sampling from memory underlies action selection during abstract decision-making. *Curr. Biol.* 32, 1–12
68. Gilboa, A. and Marlatte, H. (2017) Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* 21, 618–631
69. Ghosh, V.E. and Gilboa, A. (2014) What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia* 53, 104–114
70. Ghosh, V.E. *et al.* (2014) Schema representation in patients with ventromedial PFC lesions. *J. Neurosci.* 34, 12057–12070
71. Fellows, L.K. and Farah, M.J. (2007) The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb. Cortex* 17, 2669–2674
72. Bongioanni, A. *et al.* (2021) Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature* 591, 270–274
73. Chun, M.M. *et al.* (2011) A taxonomy of external and internal attention. *Annu. Rev. Psychol.* 62, 73–101
74. Aly, M. and Turk-Browne, N.B. (2016) Attention stabilizes representations in the human hippocampus. *Cereb. Cortex* 26, 783–796
75. Aly, M. and Turk-Browne, N.B. (2017) How hippocampal memory shapes, and is shaped by, attention. In *The Hippocampus from Cells to Systems* (Hannula, D.E. and Duff, M.C., eds), pp. 369–403, Springer International Publishing
76. Günseli, E. and Aly, M. (2020) Preparation for upcoming attentional states in the hippocampus and medial prefrontal cortex. *eLife* 9, e53191
77. Dayan, P. (1993) Improving generalization for temporal difference learning: the successor representation. *Neural Comput.* 5, 613–624
78. Momennejad, I. *et al.* (2017) The successor representation in human reinforcement learning. *Nat. Hum. Behav.* 1, 680
79. Dayan, P. (2022) Metacognitive information theory. *PsyArXiv* Published online September 18, 2022. https://doi.org/10.31234/osf.io/azujr
80. Cortese, A. *et al.* (2020) Unconscious reinforcement learning of hidden brain states supported by confidence. *Nat. Commun.* 11, 4429
81. Folke, T. *et al.* (2016) Explicit representation of confidence informs future value-based decisions. *Nat. Hum. Behav.* 1, 0002
82. Fleming, S. *et al.* (2018) Neural mediators of changes of mind about perceptual decisions. *Nat. Neurosci.* 21, 617–624
83. De Martino, B. *et al.* (2012) Confidence in value-based choice. *Nat. Neurosci.* 16, 105–110
84. Sanders, J.I. *et al.* (2016) Signatures of a statistical computation in the human sense of confidence. *Neuron* 90, 499–506
85. Lebreton, M. *et al.* (2015) Automatic integration of confidence in the brain valuation signal. *Nat. Neurosci.* 18, 1159–1167
86. De Martino, B. *et al.* (2017) Social information is integrated into value and confidence judgments according to its reliability. *J. Neurosci.* 37, 6066–6074
87. Brus, J. *et al.* (2021) Sources of confidence in value-based choice. *Nat. Commun.* 12, 7337
88. Lak, A. *et al.* (2014) Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* 21, 617–624
89. Knudsen, E.B. and Wallis, J.D. (2021) Hippocampal neurons construct a map of an abstract value space. *Cell* 184, 1–11
90. Mante, V. *et al.* (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84
91. Panichello, M.F. and Buschman, T.J. (2021) Shared mechanisms underlie the control of working memory and attention. *Nature* 592, 601–605
92. Schuck, N.W. *et al.* (2015) Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron* 86, 331–340
93. Mazor, M. *et al.* (2020) Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *eLife* 9, e53900
94. Lebreton, M. *et al.* (2009) An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* 64, 431–439
95. Lee, S.W. *et al.* (2014) Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699
96. Mansouri, F.A. *et al.* (2017) Managing competing goals – a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* 18, 645–657
97. Rangel, A. *et al.* (2008) A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556
98. Ciaramelli, E. *et al.* (2021) The role of ventromedial prefrontal cortex in reward valuation and future thinking during intertemporal choice. *eLife* 10, e67387

99. Tse, D. *et al.* (2011) Schema-dependent gene activation and memory encoding in neocortex. *Science* 333, 891–895
100. Gläscher, J. *et al.* (2012) Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14681–14686
101. Yin, S. *et al.* (2021) Ventromedial prefrontal cortex drives the prioritization of self-associated stimuli in working memory. *J. Neurosci.* 41, 2012–2023
102. Knudsen, E.B. and Wallis, J.D. (2022) Taking stock of value in the orbitofrontal cortex. *Nat. Rev. Neurosci.* 23, 428–438
103. Park, S.A. *et al.* (2020) Map making: constructing, combining, and inferring on abstract cognitive maps. *Neuron* 107, 1226–1238
104. Hayden, B.Y. and Niv, Y. (2021) The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behav. Neurosci.* 135, 192–201
105. Juechems, K. and Summerfield, C. (2019) Where does value come from? *Trends Cogn. Sci.* 23, 836–850
106. Keramati, M. and Gutkin, B. (2011) A reinforcement learning theory for homeostatic regulation. In *Advances in Neural Information Processing Systems*, MIT Press
107. Keramati, M. and Gutkin, B. (2014) Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife* 3, e04811
108. Sutton, R.S. and Barto, A.G. (1998) Reinforcement Learning: An Introduction. MIT Press
109. Schultz, W. *et al.* (1997) A neural substrate of prediction and reward. *Science* 275, 1593–1599
110. O'Doherty, J. *et al.* (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337
111. Ito, M. and Doya, K. (2009) Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J. Neurosci.* 29, 9861–9874
112. Hampshire, J.B. and Waibel, A. (1992) The Meta-Pi network: building distributed knowledge representations for robust multisource pattern recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 751–769
113. Jacobs, R.A. *et al.* (1991) Adaptive mixtures of local experts. *Neural Comput.* 3, 79–87
114. Haruno, M. *et al.* (2001) Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220
115. Sugimoto, N. *et al.* (2012) MOSAIC for multiple-reward environments. *Neural Comput.* 24, 577–606
116. Liu, C. *et al.* (2015) Multiobjective reinforcement learning: a comprehensive overview. *IEEE Trans. Syst. Man Cybern.* 45, 385–398
117. Hayes, C.F. *et al.* (2022) A practical guide to multi-objective reinforcement learning and planning. *Auton. Agent. Multi. Agent. Syst.* 36, 26
118. Yang, R. *et al.* (2019) A generalized algorithm for multi-objective reinforcement learning and policy adaptation. *arXiv* Published online November 6, 2019. https://doi.org/10.48550/arXiv.1908.08342
119. Tomov, M.S. *et al.* (2021) Multi-task reinforcement learning in humans. *Nat. Hum. Behav.* 5, 764–773
120. Sharp, P.B. *et al.* (2022) Humans perseverate on punishment avoidance goals in multigoal reinforcement learning. *eLife* 11, e74402
121. Cela-Conde, C.J. *et al.* (2004) Activation of the prefrontal cortex in the human visual aesthetic perception. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6321–6325
122. Kant, I. (1987) Critique of Judgment. Hackett Publishing
123. Goldman, A.H. (2018) Aesthetic Value. Routledge
124. Zeki, S. (2002) Inner vision: an exploration of art and the brain. *J. Aesthet. Art Critic.* 60, 365–366
125. Biederman, I. and Vessel, E. (2006) Perceptual pleasure and the brain: a novel theory explains why the brain craves information and seeks it through the senses. *Am. Sci.* 94, 247–253
126. Iigaya, K. *et al.* (2021) Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nat. Hum. Behav.* 5, 743–755
127. Sarafyazd, M. and Jazayeri, M. (2019) Hierarchical reasoning by neural circuits in the frontal cortex. *Science* 364, eaav8911
128. Collins, A. and Koechlin, E. (2012) Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* 10, e1001293
129. Donoso, M. *et al.* (2014) Foundations of human reasoning in the prefrontal cortex. *Science* 344, 1481–1486
130. Ortoleva, P. (2012) Modeling the change of paradigm: non-Bayesian reactions to unexpected news. *Am. Econ. Rev.* 102, 2410–2436
131. Botvinick, M. *et al.* (2019) Reinforcement learning, fast and slow. *Trends Cogn. Sci.* 23, 408–422
132. Doya, K. (2002) Metalearning and neuromodulation. *Neural Netw.* 15, 495–506
133. Wang, J.X. *et al.* (2018) Prefrontal cortex as a meta-reinforcement learning system. *Nat. Neurosci.* 21, 860–868
134. Schweighofer, N. and Doya, K. (2003) Meta-learning in reinforcement learning. *Neural Netw.* 16, 5–9
135. Doya, K. *et al.* (2002) Multiple model-based reinforcement learning. *Neural Comput.* 14, 1347–1369
136. Cohen, Y. and Schneidman, E. (2013) High-order feature-based mixture models of classification learning predict individual learning curves and enable personalized teaching. *Proc. Natl. Acad. Sci. U. S. A.* 110, 684–689
137. Kawato, M. and Cortese, A. (2021) From internal models toward metacognitive AI. *Biol. Cybern.* 115, 415–430