Title

The affective gradient hypothesis: an affect-centered account of motivated behavior

Author

Shenhav, Amitai

Publication Date

2024-09-01

DOI

10.1016/j.tics.2024.08.003

The affective gradient hypothesis:
An affect-centered account of motivated behavior

Amitai Shenhav
Department of Psychology
Helen Wills Neuroscience Institute
University of California, Berkeley

In press, *Trends in Cognitive Sciences*

*Correspondence: amitai@berkeley.edu

Keywords*:* motivation; decision-making; value; goal-directed behavior; emotion; cognitive control; self-regulation

Abstract

Everyone agrees that feelings and actions are intertwined, but cannot agree how. According to dominant models, actions are directed by estimates of value, and these values shape or are shaped by affect. I propose instead that affect is the only form of value that drives actions. Our mind constantly represents potential future states and how they would make us feel. These states collectively form a gradient reflecting feelings we could experience depending on actions we take. Motivated behavior reflects the process of traversing this affective gradient, towards desirable states and away from undesirable ones. This Affective Gradient Hypothesis solves the puzzle of where values and goals come from, and offers a parsimonious account of apparent conflicts between emotion and cognition.

I can't stop this feeling…

More often than not, it seems that the answer to why we did or didn't do something carries a **feeling** at its core [1] (see Glossary). We choose a restaurant, movie, or vacation destination because of how much we think we would enjoy it relative to others. We go to the gym or persist with a mentally demanding task because doing so will make us feel good (at least in the long term) and/or because failing to do so would make us feel bad. It can seem at times as though feelings carry us through our moment-to-moment decisions about what to think and how to act as we move through our day and consider aspects of our environment, both observed (e.g., the behavior and expressions of those around us) and imagined (e.g., how they would react to our words and actions).

And yet, when contemporary models seek to explain what gives rise to a set of actions, feelings often fade into the background. Instead, "cold" estimates of **value** take center stage. Value-based accounts have sought to explain how we direct our thoughts [2,3] and actions [4,5], whether deliberately (**goal-directed** behavior) or impulsively (**Pavlovian** behavior) [6–8]. I will refer to these collectively as **motivated behaviors**, distinguishing them primarily from behaviors that are determined through "value-free" (e.g., habitized) processes [9]. I will argue that prevailing accounts of motivated behavior are inside out –feelings not only play a bigger role than previously acknowledged, but that they may, *on their own*, be sufficient to explain the causes of all motivated behavior.

In the shadow of (a different sort of) value: Feeling's place in theories of motivated behavior

For over a hundred years, dominant models of motivation and decision-making have built on insights from economic theory that describe how a person determines the best course of action in a given situation [5,10]. At their core, these models propose that people integrate information about potential future outcomes (e.g., lunch options or job offers) to determine a unitary scalar estimate of the expected reward (or utility) associated with each, and then choose actions that will maximize this value. Over the years, researchers have proposed multiple ways in which feelings (under the rubric of **affect** and/or **emotion**) might intersect with the core of this framework (Fig. 1).

At one extreme is the proposal that feelings are merely epiphenomena of value (Fig. 1A). Under this account, value is derived from aspects of one's internal and external environment, and particular configurations or transformations of these values produce experiences of certain feeling states. For instance, some accounts propose that regret can reflect counterfactual representations of foregone value [11]; that happiness can reflect aggregated prediction errors [12]; and/or that mood states can reflect the aggregate expected reward in one's environment [13,14].

Another possibility is that value is shaped by feelings. One version of this account proposes that values are learned from (i.e., originate in) the feelings that arise while experiencing a given outcome (**experienced utility**) – for instance, our enjoyment or distaste when eating at a particular restaurant gets transformed into a value that in turn drives decision-making [15–17] (Fig. 1B).

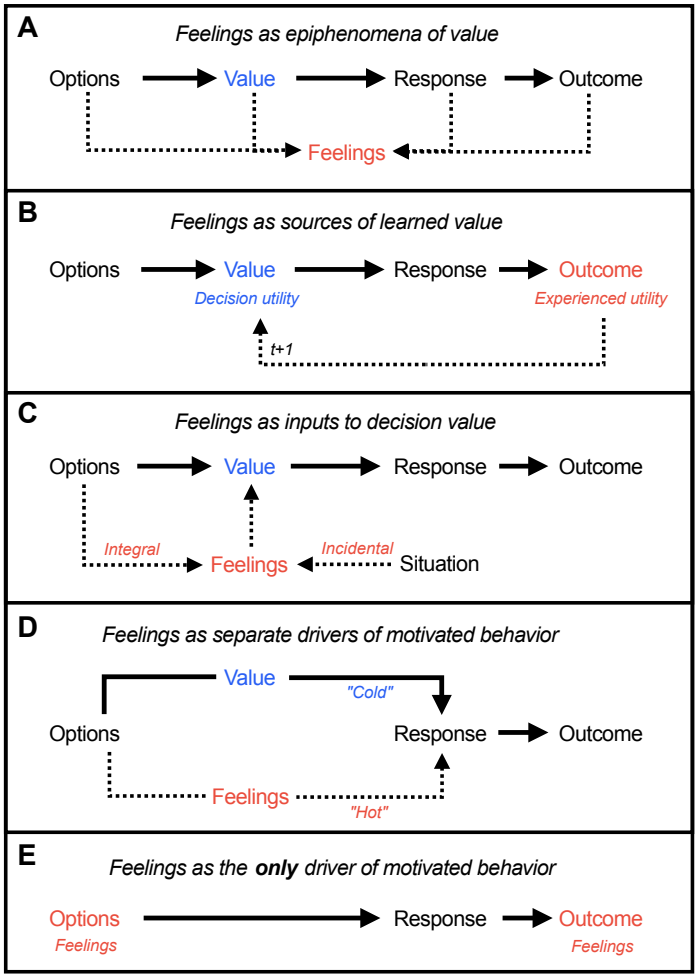

Figure 1. The role of feelings in decision-making models. Existing models of decision-making propose several potential roles for feelings (e.g., affect, emotion, mood): **(A)** as a read-out of ongoing evaluations; **(B)** as learning signals that are abstracted into an estimate of value or utility; **(C)** as relevant (integral) or irrelevant (incidental) inputs to estimates of decision value; and **(D)** as drivers of a separate ("hot" or Pavlovian) system for action control. These roles are illustrated individually, but contemporary accounts often combine several of these, in some cases differentiated by the types of feelings evoked (e.g., shorter- vs. longer-lasting). **(E)** Independent of their specific role(s) for feelings, all of these accounts maintain that a separate system exists to evaluate potential outcomes in a way that abstracts from or sidesteps feelings. The current account proposes that feelings alone (in the form of affect; cf. experienced utility) may be sufficient to account for all motivated behaviors, without the need for a separate system for "colder"/more goal-directed evaluations.

Other accounts propose that decision values are informed by feelings at the time of a decision, through one of two routes: (a) deliberate consideration of feelings about possible future outcomes (e.g., risks) [18–23] or (b) ongoing feelings (e.g., mood states) that co-occur with a decision [24,25] (Fig. 1C).

Feelings have also been proposed to drive a more primitive system that exerts direct control over actions (e.g., a Pavlovian or "hot" system), separately and/or in competition with a system that

determines action based on more deliberate evaluations of future outcomes (e.g., goal-directed or "cold" systems) [7,8,26–28] (Fig. 1D). For instance, cravings may drive reward-related impulses that hijack a decision towards an unhealthy food rather than a healthy one.

These accounts all share the assumption that there is a form of value that drives behavior independent of feelings. However, in doing so, they force us to contend with deeper puzzles that remain far from resolved. If value can be determined independently of feelings (e.g., Fig. 1A, D), what other ways do we have to assess value? And how are these two sets of values weighed against one another to determine our actions? If values are derived from feelings (e.g., Fig. 1A-C), what transformations take place that allow us to abstract feelings about everything from meals to movies to partners to vacations into a common value representation (while remaining robust to changing goals and contexts [29,30])? These puzzles may yet be solvable, but they also create an opening for a radical alternative – that feelings may, alone, be sufficient to account for the same range of behaviors as are currently driven by "cold" values (Fig. 1E).

Nothing more than feelings?

The idea that feelings may be the only form of value is not entirely new (see especially work by Zajonc [1], De Sousa [31], and Damasio [32]), but it has been largely absent from formal accounts of motivation and decision-making. There are at least two reasons why such models have felt the need to distinguish – and center on – a valuation system separate from feelings, both of which serve as barriers to adopting a feelings-only account.

One reason for skepticism: Feelings as relatively infrequent

First, feelings-related constructs (e.g., affect, emotion, mood) are often used to refer to subjective states that are particularly salient (i.e., "non-neutral"), extended in time, and/or involve an aroused bodily state [22,33]. Feelings like these are not constant occurrences for most people, and are therefore limited in the range of behavior they can account for. This seems to call for a supplemental system to account for "cold" evaluations that drive behavior when we are not in these more extreme states (e.g., mundane decisions/tasks) [26,34].

This barrier can be overcome by focusing our definition of feelings on an established use of the term *affect* (or core affect) [35,36]. At the broadest level, affect refers to a feeling that varies continuously along multiple dimensions (e.g., valence and arousal), including ones that might be referred to as 'neutral' and small variations around those (e.g., feeling moderately positive or negative about one's lunch). Affect is not directly synonymous with categorizable emotions like sadness, anger, and schadenfreude, but likely underpins them. Affect is also not synonymous with specific bodily states (e.g., changes in heart rate or gastric motility), but may abstract over these [37,38]. Further, affect can refer to feelings that occur over very brief periods (e.g., a fleeting reminder of a funny joke you heard earlier, the dreaded possibility that you forgot to bring your coffee to work with you, or the premonition of an impending collision between one's toddler and

the nearest coffee table). Finally, and critically, affect is not unitary within any particular moment – we can hold multiple competing feelings at once (or in rapid alternation), and these can be tied to multiple sources (e.g., pain in your joints simultaneous with enjoyment of a meal) and/or to a common source (e.g., feeling both excited and anxious about performing a task) [39,40].

Affect has all of the necessary characteristics to serve as an all-purpose guide to motivated behavior. It is a fundamental property of our phenomenal experience [1,35,36], emerging at the earliest stages of development [41] and accessible to us throughout our waking (and dreaming) life [42]. Most importantly, affect also has an evaluative quality [35,40,43] – we have a sense of how something feels to us relative to a feeling we would like to be having in that moment or in the future (e.g., whether it is something we would like to approach or avoid) [44]. This value signal is analogous to what has been referred to as experienced or hedonic utility [16,17].

Another reason for skepticism: Feelings as involuntary

A second reason that feelings have been previously viewed as insufficient for holding together models of motivated behavior stems from a view that they are *involuntary* reactions to aspects of our environment. To the extent this is the case, it is easier to envision feelings as drivers of instinctive actions (e.g., impulses) than of complex forms of deliberation and planning (e.g., writing a grant). It is hard to imagine achieving the level of flexibility and goal-directedness required of the latter tasks with only a reflexive valuation process.

Adopting the definition of affect above, I will argue that this additional barrier can be overcome if, instead of viewing affect as arising through reflex rather than deliberation, we view affect as a *property* of thoughts that can arise through either reflexive or deliberative means. To elaborate on this argument – and pave a path towards reconciling an affect-centered viewpoint with prevailing models of learning and decision-making – I will begin by describing how affect can be understood as a general property of our internal model of potential future states of our environment. I will then go on to describe how motivated behavior can emerge naturally from optimizing behavior along this dynamically changing **state space**, without requiring any additional value representations.

An affect-infused state space

It is now well-established that humans and other animals maintain an internal model of their environment [45,46], consisting of rich and structured associative maps relating potential locations/contexts, objects, episodes, and concepts [47–49]. These maps enable us to revisit past states of our environment and project into potential future states [50–52], a process that can transpire voluntarily (e.g., through directed search) or involuntarily (e.g., through an automatic spreading of associations; cf. priming) [50,53–55].

These state-space representations form a critical interface between perception, memory, and action. I will argue that, in so doing, they enable action to be dynamically influenced by affective content

embedded in these state representations. To ground this argument, I begin by extrapolating from existing work the basic principles underlying such an embedded structure (Fig. 2):

1) **There are affective qualities or *features* to every state that a person can represent.** The state of eating a meal, being given negative feedback, or paying one's bills all have affective features that carry a specific identity (e.g., the taste of a particular food) and scalar intensities along a limited set of dimensions (e.g., valence and arousal) (Fig. 2A). In other words, affect is both multivariate and – similar to perceptual features like color and depth [31,56] – evoked by any stimulus or context that is brought to mind.
2) **Affective features can be evoked by experienced, recalled, or imagined states.** Affective features are evoked while a person is in a relevant state (e.g., eating a meal) and when those states are brought to mind through recollection (e.g. recalling a recent meal) or prospection (e.g., imagining having that meal at a restaurant tonight) (cf. [18,44]; Fig. 2B).
3) **Affective features can be evoked in a "bottom-up" or "top-down" fashion.** A particular state can be brought to mind via an external prompt (e.g., a poster for a particular restaurant) and/or subsequent spreading of associations (e.g., being reminded of meals you've eaten at this restaurant and other similar restaurants), thus evoking the affect tied to that situation in a "bottom-up" fashion [57–59]. These same states can also be brought to mind through forms of directed search (e.g., considering potential dinner options), constituting a "top-down" route to accessing the same affective experience (cf. [7,51,60,61]) (Fig. 2C).
4) **Multiple affective features can be evoked in parallel**. A person can (near-)simultaneously bring to mind states relevant, for instance, to (a) an immediate decision (e.g., different approaches to writing a section of a grant); (b) their overall task (e.g., whether to try to meet the upcoming grant deadline); (c) other potential activities (e.g., checking their e-mail); and (d) basic survival instincts (e.g., hunger, pain from a recent injury) (cf. [39]) (Fig. 2D).
5) **The salience of an affective feature scales with the salience of its associated state**. Anything that make a given state more accessible also increases the salience of associated affective features (cf. [17,19,62]). Conversely, variables known to decrease accessibility (or render a state more "psychologically distant" [63]) should similarly weaken the relevant affective experience – this includes factors that make a given outcome seem improbable, spatially or temporally distant, or otherwise unlikely to impact oneself. This should be true independent of the affective content, in that more vivid outcomes should produce more salient affective experiences whether neutral, intensely positive, or intensely negative (cf. [64,65]). For instance, the grant writer might be motivated by the possibility of this grant being funded and/or the possibility of failing to secure *any* grant funding, but the affective features of these states may be less salient than those of states they perceive as more immediate and likely, such as the possibilities of successfully submitting the grant or missing the deadline (Fig. 2E).

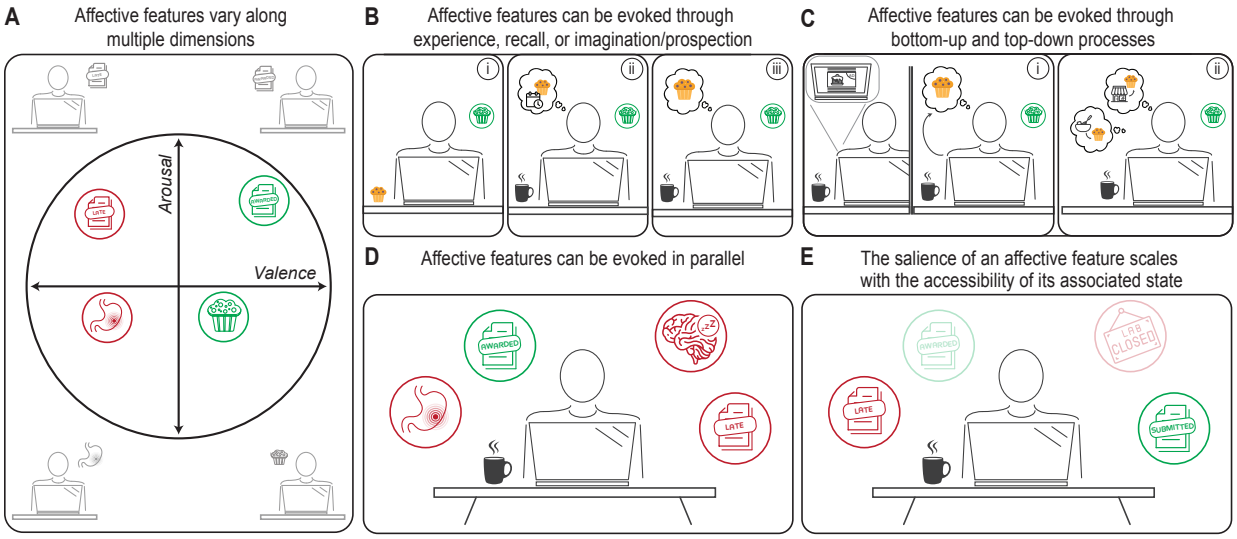

Figure 2. Affect as a feature within an internal state space. According to the current account, every internally represented state (e.g., episode) carries affective features. **A)** Each of these states (e.g., current hunger level, eating a muffin, submitting a grant late, having a grant awarded) are represented along a limited set of affective dimensions (e.g., valence and arousal). **B)** States can be brought to mind through experience, recall, or prospection (e.g., current, past, or future meal), and in each case evoke the associated affective features (e.g., positive feelings towards eating the food in question). **C)** States and associated features can be brought to mind through bottom-up cueing (e.g., an image on the screen) or directed search (e.g., choosing a lunch venue). **D)** A person can have in mind affective features associated with multiple states at the same time (or in rapid alternation) (e.g., current feelings of hunger and mental fatigue, future outcomes related to success or failure at the current task). **E)** States that are most accessible and/or vivid (e.g., ones that are perceived to be more immediate or likely) will have more salient affective features.

Traversing an affective gradient: From affect to motivation

So far, I've proposed that affective content is embedded into representations of current and potential future states, and that affective experiences are therefore constantly generated to varying degrees by all states under consideration. Thus, at any given time, a person can represent feelings associated with, among other things, options they are evaluating; consequences of persevering with or disengaging from their current task; and other potential tasks they would like to pursue.

The landscape of affective features being held in mind at any given moment can also serve as a guide for how to adjust behavior in that moment. States that are expected to improve one's affect serve as attractive landmarks (those to be reached), whereas states that are expected to worsen one's affect serve as repulsive landmarks (to be avoided). The actions afforded by one's current environment can each be described in terms of their relationship with these states: to what extent does taking that action increase the likelihood of reaching more positive states and/or avoiding more negative states (Fig. 3A)? Each action can thus be described as occupying a location along a multidimensional **gradient** (Box 1), and this location identifies the expected affective consequences of taking that action (Fig. 3B). For instance, checking your e-mail reduces aversive

uncertainty and/or increases the potential for positive surprise arising from a recent notification [66,67]; focusing on the task you've just started reduces the likelihood that you will miss a deadline or have to stay late; and stopping to eat your lunch reduces growing hunger and momentarily increases feelings of satisfaction and enjoyment.

The affective gradient hypothesis (AGH) proposes that the gradient formed out of currently accessible states acts as an **objective function** for guiding motivated behavior, and that actions and control states are actively and dynamically mobilized in service of this objective function. In other words, expectations of potential future affect promote actions and control states that minimize expected negative affect and maximize expected positive affect, and suppress behavior that achieves the opposite ends.

AGH builds on previous gradient-like accounts that have been used to describe the push-pull influence of potential reward and punishment on approach and avoidance behavior [68–70] (Box 1). For instance, an animal can be simultaneously drawn towards a positively-valenced outcome (e.g., food) and away from a negatively-valenced outcome (e.g., shock). These drivers of approach and avoidance have traditionally been characterized as reflecting primitive (Pavlovian) forms of value-based control, too rigid and reflexive to account for the goal-directedness that characterizes most human behavior [6]. However, a more expansive view of when and how affect is generated (Fig. 2) – one that ties affective features to states accessed either automatically or deliberately – paves a path for a Pavlovian-like form of value (i.e., affective features in my model) to scaffold complex goal-directed behaviors.

Consider a person faced with a grant deadline. They might have in mind potential consequences of submitting or failing to submit a grant, each of these a state carrying corresponding affective features (e.g., feeling good about submitting) and further associated states (e.g., being awarded the grant) (Fig. 3A). Traversing this affective gradient entails pursuing a particular course of action (e.g., writing the grant) (Fig. 3B). The persistent availability of these and other affective consequences further scaffolds the rest of the writing process – to make progress towards submitting the grant (e.g., to avoid negative outcomes), individual parts need to be completed, and completing each of those entails evaluation of other states (e.g., the viability of different sub-aims).

Canonical value-based models revisited

As the previous example makes clear, AGH inverts affect's typical role in models of motivation and decision-making (Fig. 1A-D). These models typically start from the imperative to make a choice (e.g., presentation of options), and work outwards to incorporate affect into the process of evaluation and/or response selection. AGH instead centers on fluctuations in expected affect (e.g., how one would feel if they did or did not act in a certain way), and proceeds outward to determine decisions and actions. Expectations that one's affect could be improved or worsened promote actions aimed at achieving the former and/or avoiding the latter (to the extent the person perceives

such actions as being under their control [71,72]). Thus, affect does not serve to drive decisions; rather, decisions are made because they serve to optimize affect. By the same token, AGH predicts that expected affect should determine not only which options are chosen, but also whether and how a decision is made [73].

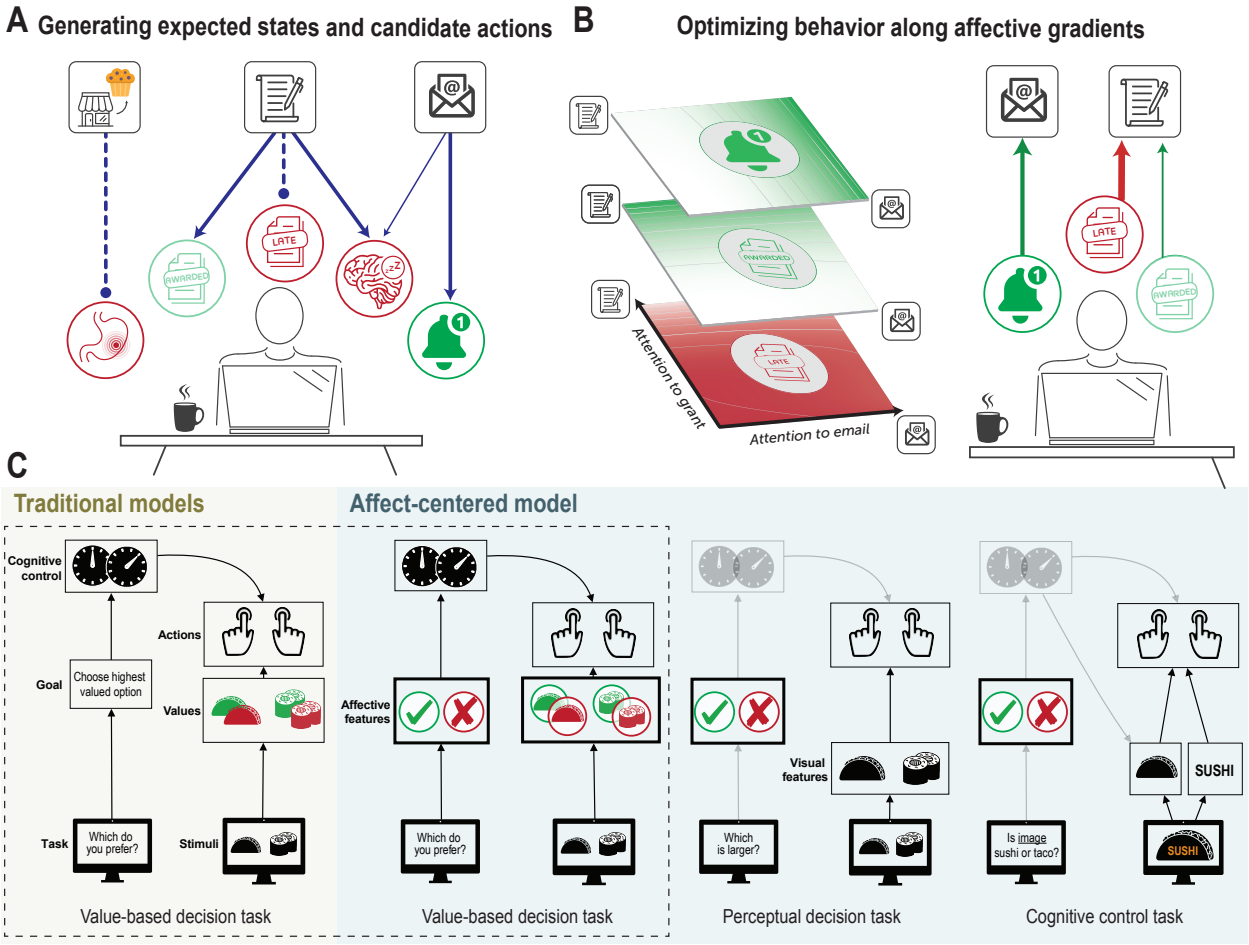

Figure 3. Expected affect motivates behavior. A) While working on a given task (e.g., grant-writing), a variety of states may become accessible, some with positive affective features (e.g. grant being awarded, uncovering the contents of an incoming e-mail) and others with negative affective features (e.g., hunger, mental fatigue, missing a deadline), each varying in its salience (Fig. 2E). Each of these may bring to mind actions (e.g., walking to store, attending to grant, attending to e-mail) that can make a given state more likely (solid arrows) or less likely (dashed arrows) to occur. **B)** The relationship between a given action (e.g., attending to grant) and a given state (e.g., grant being awarded) can be described by a gradient, with increased attention making it more likely that the person will experience the affect associated with that state (e.g., elation). Other actions (e.g., attending to e-mail) produce different gradients with respect to this and other potential consequences (e.g., feelings associated with reading the e-mail). AGH proposes that actions are optimized so as to maximize expected positive affect and minimize expected negative affect, for instance in this case by maintaining focus on the task at hand. **C)** Applying this model to a value-based decision task (e.g., whether to eat sushi or tacos), both AGH (blue) and traditional models (yellow) predicts that actions should be guided by the affect expected from obtaining a given option (e.g., based on past experiences

with each). AGH diverges from these models in how it accounts for one's engagement in the task (e.g., level of attentional focus, threshold for making a decision; see Box 1). Traditional models assume that the task serves as its own goal and participants adjust control when performance worsens (e.g., by monitoring for errors or conflict). AGH instead proposes that task engagement is directly determined by the salience of potential performance outcomes (e.g., likelihood that one would be perceived to be performing the task poorly) and how these outcomes would feel. Thus, actions (e.g., left vs. right) and control states (e.g., levels of attention and threshold) are both determined by expected affect. Rather than maintaining a task goal explicitly, a person need only represent contingencies between action and performance (e.g., task rules) and the consequences of performing well or poorly. This account generalizes across task rules, including those that focus only on visual features of stimuli (e.g., size or form) rather than their affective features.

Consider a study participant who is asked to choose which of two options they prefer. AGH's account of the focal decision is similar to common integral affect-based theories (Fig. 1C) [18–21] – participants draw on relevant affective experiences (e.g., associated with monetary loss or consumption of a given food) and choose the option that they anticipate will produce the greatest positive/least negative affect. These accounts share the prediction that, in cases like these, measures of affect will predict decision outcomes better than objective estimates of reward value (e.g., [74–76]), and choices will be shaped by biases and constraints in representing affect [44,62].

However, this shared account of the focal decision falls short of explaining why the person decides to make a choice in the first place (rather than, e.g., pressing buttons randomly or walking out of the experiment). For a given task, this gap can be filled by assuming that participants maintain an instructed goal (e.g., choose the best option), and allocate control (e.g., selective attention) in service of that goal [77] (Fig. 3C, left). AGH offers a more generalizable solution to this problem, by instead simply assuming that participants represent relevant control contingencies (e.g., likelihood that a given control configuration produces a correct or incorrect response) and the affective features of each of those outcomes (Fig. 3C, right). These affective expectations can promote persistence and/or increased effort when stakes increase and/or become more salient (e.g., following an error [78–80]); they can also promote effort divestment and task-switching as stakes decrease or outcomes seem more assured (e.g., easy tasks) (cf. [81]). In providing an account of the dynamics of motivated behavior that generalizes within and across tasks, AGH thus also helps resolve a deeper puzzle within research on goal-directed behavior: how do people know *which* goal to pursue at a given time (Box 2)?

Broader implications

Warming up: You're hot, then you're (not) cold

One of the oldest and most persistent dichotomies in psychology distinguishes between emotional and cognitive processes. Within models of judgment and decision-making, this distinction has manifested in a dichotomy between "hot" and "cold" systems or modes of evaluation [26,28,34]. The traditional assumption is that the "hot" system can be activated by triggering certain affective reactions (e.g., with vivid images or environmental stressors), and that doing so biases a person to

act in a way different than they would if they were performing a "cold" evaluation of those same options.

Under the current account, there are no "cold" (i.e., affect-free) evaluations; affect serves as the underpinning for all motivated behavior, whether reflexive/Pavlovian or deliberate/goal-directed. Thus, processes that were previously distinguished as hot versus cold don't reflect differences in *whether* affect is involved, but rather *how* it is involved (e.g., with varying intensities or levels of arousal). Accordingly, AGH predicts that evaluation should always entail representing the affective features of expected outcomes. Moreover, identical outcomes could share overlapping neural codes, whether accessed automatically (e.g., Pavlovian) or deliberately (e.g., goal-directed).

Recasting intrapersonal conflict: One self, many feels

Accounts of self-regulation describe conflicts that arise between one's current inclinations and those of an idealized actor, where the latter can reflect, for instance, projections to a future self (e.g., one that has health, wealth, and happiness) or other forms of social norms or moral ideals (e.g., being a good person, doing what's right) [82]. Popular (yet controversial [28,83]) accounts cast these intrapersonal conflicts in terms of competing agents (selves) within one's mind, each with its own objective function (e.g., maximizing immediate vs. long-term reward, serving personal vs. group/societal interests) [84,85].

Under AGH, such conflicts can instead simply reflect the representation of multiple competing affective consequences of one's actions (cf. **ambivalence** [39,86]) – e.g., it would feel good to enjoy this dessert, but would feel bad to be judged negatively if I gain weight; it would feel bad to sacrifice some of my money, but would feel good to think of others benefiting as a result; it would feel good to kick this habit, but would feel bad if I didn't satisfy my craving [87]. AGH predicts that the manner in which these conflicts resolve (e.g., whether self-regulation "succeeds" or "fails") will depend on the relative strength and salience of these affective representations (Box 3), rather than the strength with which a particular self is represented.

Common objectives without the need for a common currency

A standard assumption across most neuroeconomic models is that the brain integrates different sources of value into a **common currency**, which it uses to compare the values of different options and select the best one [5,88,89]. Under AGH, the same could be true – e.g., affective features could be integrated and compared along a composite measure of valence and intensity (cf. [20,43]) – but it need not be. Instead, motivated behavior can arise from affective features influencing relevant actions and control states directly and in parallel (Box 1, Fig. 3) (cf. [90,91]).

By obviating the need for a common currency, AGH can also avoid the significant theoretical and empirical challenges these accounts face [92,93], for instance in explaining how people compare option values that differ in the dimensions along which they are valued (cf. Box 2). With its focus

on state-specific affective features, AGH arguably magnifies this concern (e.g., it's hard to conjure up a feeling associated with the *difference* between having sushior tacos for dinner), and thus reinforces the possibility that potential states can only be compared with one another in terms of their influence on potential actions.

If affective features can't be compared directly, AGH goes further to predict that people will also have trouble representing counterfactuals to a decision (e.g., how much they might regret taking a certain action) strictly in terms of the marginal benefits of the chosen option relative to the best alternative (as prescribed by economic theory [94]); instead, these evaluations might be shaped by the values of each option being considered (cf. [95]). Similarly, neural signals that appear to reflect scalar comparisons (e.g., the value of one option relative to others [88,96]) would have to instead reflect representations of individual options (e.g., positive affect associated with obtaining a given option, negative affect associated with sacrificing alternatives) or metacognitive (e.g., task-level) representations (e.g., likelihood and consequences of choosing in/correctly [77,97,98], Fig. 3C).

Concluding remarks

The account I have offered forces significant revisions to dominant views across psychology and neuroscience. While this may give pause in accepting the underlying premise, it's important to recall the current state of affairs, absent this alternative: Researchers agree that affect is a persistent feature of our experience and plays critical roles in shaping evaluations and actions – sometimes intentionally [18,19,22], sometimes less so [24,25] – and have sought to distinguish these roles from a separate mode of decision-making that is driven by an affect-less form of value [1,26]. But in doing so, we have ended up with a mobius strip of unresolved transformations and interactions (Fig. 1). Research into goal-directed behavior more generally has been built on a foundation of goals that each effectively serve the role of middle managers (providing directions in the service of their designated aims) but without a clearly articulated objective function ("CEO") to determine how goals are prioritized relative to one another in a given moment (e.g., should I be more concerned with completing this task or satisfying my hunger?), nor even clear and consistent boundaries to define those goals (e.g., is my current goal to complete this task, to perform a certain way, to be a team player, or to avoid thinking about food?). Meanwhile, research across these areas has been largely unmoored from the rich mental life we occupy outside of these well-studied tasks, one that includes thoughts and feelings that come to mind both spontaneously (e.g., mind-wandering, rumination) and deliberately (e.g., reflection, imagination) [42,50].

I have proposed that affect can be construed as an evaluative feature of one's mental state space, and that in this role it can serve as a rich source of experiences of one's immediate, past, and future environments, while at the same time serve to motivate thoughts and actions towards achieving better affective experiences and avoiding worse ones. On this account, affect is not an input to or output of some other form of value. Rather, affect is the *only* form of value driving behavior, one that is phenomenologically accessible (cf. experienced utility) and multitudinous rather than

abstracted and unitary. Furthermore, on this account, goals don't need to be prioritized relative to one another. Rather, goals *reflect* the collective affective priorities of the individual at a given moment in time (Box 2, Fig. 3; cf. [31]).

This account lays the groundwork for future work aimed at building on this framework, and points toward fruitful avenues for doing so. Methodologically, it suggests that researchers would benefit from assaying affective features and affordances that may be available to a participant throughout an experiment [99,100], including those unrelated to the task itself [50]. These assays can borrow from any element of the standard toolbox for measuring affect (e.g., self-report, peripheral physiology, neuroimaging) [101–103], as well as recent advances in decoding neural representations of states [45,48] and their associated affective features [104–106]. Computational models should similarly seek to account for potential actions the person may consider outside the immediate task, including forms of "meta-control" [77] and alternative tasks or activities [107]. More broadly, AGH suggests that understanding within- and between-person variability in processes like motivation and self-regulation – including in the context of development and psychopathology – will ultimately require understanding how people vary in the states they bring to mind in support of these functions (Box 3).

Important gaps remain in this account (see Outstanding Questions), and addressing these represents its own challenge. However, the size and scope of existing gaps suggests that the time is ripe to give this new, affect-centered, approach a try. Perhaps, just to see how it feels.

Box 1. The role of gradients in motivation and other forms of optimization

Gradients have long played a central role in research on motivation. Early observations that animals sped up as they neared the end of a learned maze formed the basis of Hull's proposal that motivation increased progressively with decreasing distance to one's goal (the "goal-gradient hypothesis") [108]. This prediction has since been generalized to humans pursuing goals over both short and long timescales [109,110]. Further research examined distinct gradients that emerged as a function of approach-related motivation (towards desirable outcomes) versus avoidance-related motivation (away from undesirable outcomes), which can overlay on top of each other to promote the same goal (e.g., working hard on a term paper to get a good grade and avoid letting down one's parents) or different goals (e.g., wanting to get a good grade but not wanting to exert effort) [68,70,111] (Fig. I). These approach and avoidance gradients described competing forces that could promote dynamic changes in behavior and physiology in the service of each goal, and corresponding affective states related to goal pursuit (e.g., expectations of success and failure) and goal conflict (e.g., boredom, fatigue) [70,112,113].

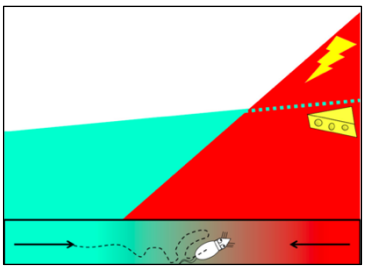

Fig. I. Approach and avoidance gradients. These describe directional influences of appetitive and aversive outcomes (e.g., cheese vs. shock) on action, as a function of distance from outcome state. Adapted from McNaughton et al. [70].

Gradients also play a critical role across fields like computer science, statistics, and engineering, where they describe the direction to move in a multidimensional space to optimize for a particular objective. Research on motivation and decision-making has increasingly drawn on these optimization approaches by characterizing action selection as a process that seeks to optimize one's future expected reward [4,114], including to describe how approach and avoidance gradients can be integrated to select among several choice options [20]. AGH builds on this work in two ways. First, drawing on research on motor and cognitive control, it assumes that the output space is multivariate, meaning that people can simultaneously evaluate multiple potential actions and control states to identify joint configurations across these (e.g., varying levels of flexion/tension of different muscle groups, varying levels of attention to different aspects of their environment) [115] (Fig. II). Second, it assumes that the objective function is also multivariate, consisting of a heterogeneous set of currently accessible affective features (e.g., distinct consequences of an action), each encoding a vector of scalar values (e.g., valence, arousal) (cf. [116,117]). Thus, optimization entails a many-to-many mapping between affective features and the potential actions that increase or decrease the likelihood of reaching

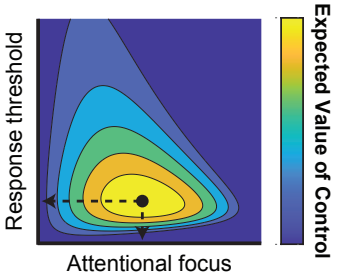

Fig. II. Optimizing multivariate control. To optimize cognitive control, previous work has generated scalar estimate of the value of control across each combination of control settings, including different levels of attentional focus (reflected in rates of evidence accumulation) and different thresholds for responding based on available evidence. The optimal control configuration (dashed arrows) can be identified by the maximal point within this gradient (black circle). Adapted from Leng et al. [97].

their corresponding states, collectively achieving a form of *multi-objective* optimization [118] (see Box 2).

Box 2. Affect as a solution to the "homungoalus" problem

Early models of cognitive control described how controllers (e.g., feature-specific attention) could intervene on stimulus-driven processes to improve performance, but failed to describe how the controllers knew that control should be enacted (Fig. I, left). This "homunculus problem" was subsequently resolved by proposing that the brain could monitor for particular sources of information (e.g., errors, conflict) to determine when and how much control is warranted, based on deviations from their goal [2,115,119,120]. However, these models – along with nearly all other models of cognition – still rely on a clear definition of the agent's current goals. This is straightforward enough when goals are instructed (e.g., attend to a word's color) or defined by a well-constrained reward function (e.g., points in an Atari game), but is much harder to resolve when broadening out to most real-world examples. How does the person writing a grant know that this is the right goal to have, and/or when they should shift priorities to other tasks, or to instead take a phone call, eat a snack, or go to the gym? This can theoretically be resolved by integrating expected rewards across all putative goals, but it is far from clear how this is done, particularly given that rewards are themselves known to depend on goals [29,30]. Thus, we are left with a "homungoalus" problem (Fig. I, right) – we know more about how thoughts and actions are coordinated once a goal is selected than we do about the ultimate objective function that determines which goals are selected, when, and for how long.

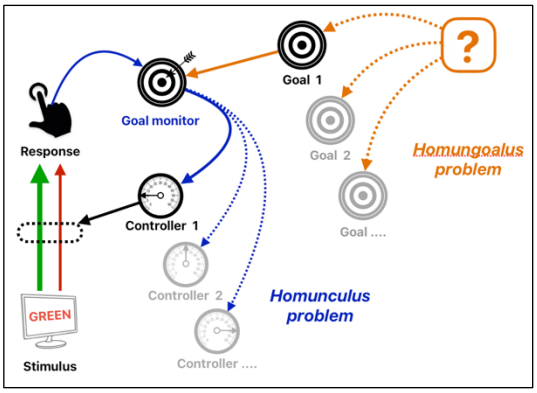

Fig. I. The "Homungoalus" Problem. Cognitive control requires a guide (Homunculus Problem), which can be achieved by monitoring performance relative to a goal. But this still leaves open the question of how goals are guided/prioritized.

One way out of this problem is to assume that people are always optimizing many goals (e.g., metabolic, social, intrinsic, etc.) [113,116,121,122], but this requires a clear definition of these discrete goal dimensions and how they are prioritized relative to one another. This can be tractable if there is a clear hierarchy and/or a common currency linking these goal values, but putative hierarchies seems potentially violable (e.g., hunger and pain can be withstood to achieve social goals) and conversion into a common currency has its own challenges (see Broader Implications). AGH solves this problem by instead removing goals from their traditional explanatory role in coordinating behavior. Under this account, behavior is dynamically reconfigured based on the affective associations that are currently accessible (e.g., related to failing to complete a task or committing a social faux pas). Goals can be viewed as an *emergent property* of this process – a person may continue to pursue their current task because they anticipate that doing otherwise would lead to greater negative affect, but soon after switch to another task because the associated increase in positive affect becomes more salient. This accounts for dynamics within and between tasks (Fig. 3C), while also capturing influences of task-*independent* objectives on moment-to-moment behavior (e.g., attending to a nearby conversation, adjusting posture and facial expressions to satisfy social conventions).

Box 3: Variability in motivated behavior within and across individuals

People vary considerably in their motivation to engage in various acts, in ways that can be maladaptive. At one extreme are cases where one lacks the motivation necessary to engage in daily activities (e.g., apathy, amotivation [123,124]). At the other extreme are behaviors that can be described as reflecting an excess of motivation for outcomes that are detrimental to their long-term health and wellbeing (e.g., drug use [87,125]). From the perspective of AGH, understanding what motivates a person to perform a certain action requires knowing what feelings they expect (a) when performing that action (e.g., effort); (b) as a result of that action (e.g., approval), (c) as a result of inaction (e.g., reprimand); and (d) as a result of other actions (e.g., foregone opportunities). For instance, someone can behave impulsively because of their positive feelings about the expected outcome (e.g., drug high), negative feelings about not achieving that outcome (e.g., frustrated craving), and/or *lack* of negative feelings about longer-term risks (e.g., addiction) (cf. [57,87,126]). Understanding how motivation differs across people, and over the lifespan, requires understanding variability in these same affective expectations.

Notably, these affective expectations each center on particular future outcomes (varying from immediate to longer term). Variability in these expected future states will not always be reflected in common instruments that assess summary estimates of a person's feelings in the moment or in the recent past (e.g., [127]). Instead, a more comprehensive approach is necessary to inventory each of the outcomes a person considers when weighing a given activity; how salient (e.g., probable) those outcomes seem; and how they would feel if a given outcome were to occur (i.e., the affective features of this outcome) (cf. measures of outcome 'expectancies'; [128]).

There is also research to suggest that people vary in the levels of affect they find most desirable [129,130]. For instance, some people strive for high-arousal positive states (e.g., elation) whereas others strive for low-arousal positive states (e.g., serenity), leading them to pursue different kinds of activities (e.g., skiing vs. hiking). It is possible that some of these **affective goals** emerge from the combination of states that are most accessible to a person (e.g., people who engaged in more vs. fewer high positive arousal activities growing up) and their associated affective features (e.g., linking arousing activities with more vs. less downside potential). Alternatively, such goals could reflect individual differences in *which* configurations of affective features (e.g., settings of valence and arousal) are most desirable. This could be conceptualized as a meta-parameter that alters the orientation of affective features with respect to potential actions (cf. [116]), effectively motivating behavior towards maximizing the desired levels of affect, whatever these may be (Fig. I).

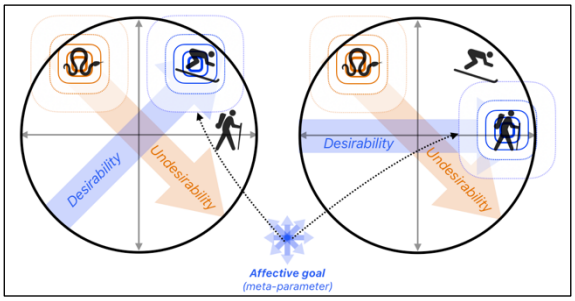

Fig. I. Affective goals as a meta-parameter on the affective gradient. Affective goals can theoretically alter the motivational impact of a given affective configuration (e.g., high-arousal positive affect), such that states with these features are either more or less attractive or repulsive. This would, in turn, promote actions that increase the likelihood of achieving this experience (e.g., skiing vs. hiking).

Glossary

- **Affect:** feelings that vary along multiple dimensions (e.g., valence, arousal) and reflect the influence of a given aspect of one's environment on the individual. Affect is distinct from (but contributes to) categorizable emotions (e.g., sadness, anger).
- **Affective goal:** the type of affective experience that a person seeks to achieve, encompassing the concepts of ideal affect (e.g., preferences for high- vs. low-arousal positive affect [129]) and emotion goals (e.g., preferences for sadness vs. anger [130]).
- **Ambivalence:** experiencing a mixture of positive and negative feelings about a given situation or action, either at the same time or in alternation with one another.
- **Common currency:** an integrative representation of value that enables different types of states or items/goods to be compared directly with one another (i.e., by converting the specific value of those states into a common currency and then comparing the common currency values with one another).
- **Emotion:** a subjective experience of one's bodily, affective, and/or motivational state that can be assigned a conceptual label (e.g., sadness, anger, elation, envy).
- **Experienced utility:** the hedonic experience associated with reaching a given outcome.
- **Feeling:** lay term encompassing changes in bodily, affective, and/or emotional state in response to an exogenous or endogenous stimulus. For a more circumscribed definition of feelings as pertains to the current model, see 'Affect.'
- **Goal-directed actions:** actions that are determined by considering the outcomes they would achieve, based on the individual's current understanding of their environment.
- **Gradient:** a set of continuous values that define the direction one needs to move within a space in order to achieve an objective or set of objectives.
- **Motivated behavior:** consists of actions that have been classified as value-based (driven by the value of expected outcomes), whether reflexively (e.g., impulses, Pavlovian actions) or through planning/deliberation.
- **Objective function:** definition of what an agent is trying to optimize (e.g., maximize) within a given environment.
- **Pavlovian actions**: evolutionarily hard-wired approach or avoidance behavior that is triggered by a learned cue-outcome association.
- **State space**: an internal representation of discrete states (e.g., locations or episodes) that a person can transition into, each defined by a set of features.
- **Value:** an estimate of reward expected from arriving in a given state (e.g., states achieved by selecting a particular option or performing a particular action), in some cases discounted by expected costs incurred to reach that state (e.g., delay or effort).

Highlights

- Models of motivated behavior suggest that feelings (e.g., affect, emotions) interact with other types of evaluations to drive action. I propose instead that affect is, on its own, sufficient to drive all motivated behavior.
- People represent potential multiple potential affective consequences at any given time. These describe the ways in which expected affect can be improved or worsened by one's actions (the affective gradient), which in turn drives dynamic adjustments in behavior.
- This affective gradient hypothesis recasts past accounts of multiple systems (e.g., hot vs. cold) and/or selves (e.g., present vs. future) driving behavior, centering instead on affect as the sole driver of behavior generated reflexively or deliberately.
- This account also helps resolve a longstanding puzzle of how goals are prioritized and maintained. It proposes that goals are an emergent property of the affective associations under consideration at a given time.

Outstanding Questions

- AGH proposes that motivation is underpinned by affective features of expected outcomes, but how are these features learned (e.g., as an integration or abstraction of current interoceptive and exteroceptive states)? To what extent do emotion categories or cognitive appraisals play additional roles in further shaping behavior)?
- Affect is argued to take the place of value in driving behaviors previously characterized as "value-based," but to what extent can elements of motivated behavior persist without affective input, including through putatively "value-free" actions (reflexes and certain forms of habits and rule-guided behavior) and/or other mechanisms that might enable continued engagement of task goals (e.g., recurrence)?
- AGH proposes that the structure of affective representations is scaffolded on the structure of one's mental state space, itself consisting of associative representations of episodic memories as well as potentially more abstract cognitive representations (e.g., concepts). The structure of that state space– including what defines the boundaries of an individual state and what factors determine the accessibility of a given state – remains poorly understood.
- AGH proposes that people optimize over many-to-many mappings between affective features and potential actions, but how this optimization is implemented remains an important open question. How are associations formed between particular affective features and particular actions, and to what extent are these tuned by evolution (e.g., relating pain-related affect to a subset of controllers), experience, and/or simulation? Is the objective for this optimization process to maximize along a particular affective axis (e.g., attain the greatest positive arousal) or to maintain a particular affective set-point? What determines a person's affective goals (cf. Box 3)?

Acknowledgments

I am grateful for invaluable feedback from Lisa Feldman Barrett and her lab, Aaron Bornstein, Peter Carruthers, Roman Feiman, Oriel FeldmanHall, Michael Frank, Ben Hayden, Michael Inzlicht, Uma Karmarkar, David Melnikoff, Gaia Molinaro, Nico Schuck, Nicholas Shea, and current and past members of my lab, including Hayley Brooks, Romy Froemer, Ivan Grahek, Xiamin Leng, Yi-Hsin Su, and Debbie Yee. I am also grateful to Alison Schreiber for helping design Figures 2 and 3. This work was supported by grants from the National Institute of Mental Health (R01MH124849) and the National Science Foundation (CAREER Award 2046111).

References

1. Zajonc, R.B. (1980) Feeling and thinking: Preferences need no inferences. *Am. Psychol.* 35, 151–175
2. Shenhav, A. *et al.* (2017) Toward a Rational and Mechanistic Account of Mental Effort. *Annu. Rev. Neurosci.* 40, 99–124
3. Gilbert, S.J. (2024) Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory. *Cognition* 247, 105783
4. Sutton, R.S. and Barto, A.G. (1998) *Reinforcement Learning: An Introduction*, MIT Press
5. Glimcher, P. (2009) Choice: towards a standard back-pocket model. *Neuroeconomics: Decision Making and the Brain* DOI: papers://7E059DFB-CCD8-4D1E-964F-6D244D60A241/Paper/p23890
6. Rangel, A. *et al.* (2008) A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556
7. van der Meer, M. *et al.* (2012) Information Processing in Decision-Making Systems. *Neuroscientist* 18, 342–359
8. Dolan, R.J. and Dayan, P. (2013) Goals and Habits in the Brain. *Neuron* 80, 312–325
9. Miller, K.J. *et al.* (2019) Habits without values. *Psychol. Rev.* 126, 292
10. Wabba, M.A. and House, R.J. (1974) Expectancy theory in work and motivation: Some logical and methodological issues. *Hum. Relat.* 27, 121–147
11. Coricelli, G. and Rustichini, A. (2010) Counterfactual thinking and emotions: regret and envy learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 241–247
12. Rutledge, R.B. *et al.* (2014) A computational and neural model of momentary subjective well-being. *Proc. Natl. Acad. Sci. U. S. A.* 111, 12252–12257
13. Eldar, E. *et al.* (2016) Mood as Representation of Momentum. *Trends Cogn. Sci.* 20, 15–24
14. Emanuel, A. and Eldar, E. (2023) Emotions as computations. *Neurosci. Biobehav. Rev.* 144, 104977
15. Rolls, E.T. and Grabenhorst, F. (2008) The orbitofrontal cortex and beyond: from affect to decision-making. *Prog. Neurobiol.* 86, 216–244
16. Berridge, K.C. and O'Doherty, J.P. (2014) From experienced utility to decision utility. In *Neuroeconomics*, pp. 335–351, Elsevier

17. Kahneman, D. and Thaler, R.H. (2006) Anomalies: Utility maximization and experienced utility. *J. Econ. Perspect.* 20, 221–234
18. Knutson, B. and Greer, S.M. (2008) Anticipatory affect: neural correlates and consequences for choice. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3771–3786
19. Loewenstein, G.F. *et al.* (2001) Risk as feelings. *Psychol. Bull.* 127, 267–286
20. Busemeyer, J.R. and Townsend, J.T. (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* 100, 432–459
21. Mellers, B.A. *et al.* (1997) Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychol. Sci.* 8, 423–429
22. Damasio, A.R. (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 351, 1413–1420
23. Roberts, I.D. and Hutcherson, C.A. (2019) Affect and Decision Making: Insights and Predictions from Computational Models. *Trends Cogn. Sci.* 23, 602–614
24. Lerner, J.S. *et al.* (2004) Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychol. Sci.* 15, 337–341
25. Schwarz, N. and Clore, G.L. (2003) Mood as information: 20 years later. *Psychol. Inq.* 14, 296–303
26. Weber, E.U. and Johnson, E.J. (2009) Mindful Judgment and Decision Making. *Annu. Rev. Psychol.* 60, 53–85
27. Huys, Q.J.M. *et al.* (2012) Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* 8, e1002410
28. Phelps, E.A. *et al.* (2014) Emotion and Decision Making: Multiple Modulatory Neural Circuits. *Annu. Rev. Neurosci.* 37, 263–287
29. De Martino, B. and Cortese, A. (2023) Goals, usefulness and abstraction in value-based choice. *Trends Cogn. Sci.* 27, 65–80
30. Molinaro, G. and Collins, A.G.E. (2023) A goal-centric outlook on learning. *Trends Cogn. Sci.* DOI: 10.1016/j.tics.2023.08.011
31. De Sousa, R. (1990) *The rationality of emotion*, Bradford Books
32. Damasio, A.R. (1994) *Descartes' error : emotion, reason, and the human brain*, G.P. Putnam
33. James, W. (1948) What is emotion? 1884. In *Readings in the history of psychology*, pp. 290–303, Appleton-Century-Crofts
34. Loewenstein, G. (2005) Hot-cold empathy gaps and medical decision making. *Health Psychol.* 24, S49-56
35. Barrett, L.F. and Bliss-Moreau, E. (2009) Affect as a psychological primitive. *Adv. Exp. Soc. Psychol.* 41, 167–218
36. Jastrow, J. (1897) Outlines of psychology . By Wilhelm Wundt. Translated by Charles Hubbard Judd. Leipzig, Wilhelm Engelmann. 1897. 8vo, pp. 342; An outline of

psychology . By Edward Bradford titchener. New York, the Macmillan company. 1897. 8vo, pp. 352. *Science* 5, 882–884

37. Craig, A.D.B. (2009) How do you feel — now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70
38. Barrett, L.F. (2016) The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* DOI: 10.1093/scan/nsw154
39. Vaccaro, A.G. *et al.* (2020) Bittersweet: The neuroscience of ambivalent affect. *Perspect. Psychol. Sci.* 15, 1187–1199
40. Cacioppo, J.T. *et al.* (2014) The evaluative space model. In *Handbook of Theories of Social Psychology: Volume 1*, pp. 50–73, SAGE Publications Ltd
41. Berridge, K.C. and Kringelbach, M.L. (2008) Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* DOI: papers://7E059DFB-CCD8-4D1E-964F-6D244D60A241/Paper/p11092
42. Fox, K.C.R. *et al.* (2018) Affective neuroscience of self-generated thought. *Ann. N. Y. Acad. Sci.* 1426, 25–51
43. Cabanac, M. (1992) Pleasure: the common currency. *J. Theor. Biol.* 155, 173–200
44. Wilson, T.D. and Gilbert, D.T. (2003) Affective forecasting. *Adv. Exp. Soc. Psychol.* 35, 345–411
45. Schuck, N.W. *et al.* (2018) A State Representation for Reinforcement Learning and Decision-Making in the Orbitofrontal Cortex. In *Goal-Directed Decision Making*, pp. 259–278, Elsevier
46. Wilson, R.C. *et al.* (2014) Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* 81, 267–279
47. Bar, M. (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289
48. Whittington, J.C.R. *et al.* (2022) How to build a cognitive map. *Nat. Neurosci.* 25, 1257–1272
49. Renoult, L. *et al.* (2019) From knowing to remembering: The semantic–episodic distinction. *Trends Cogn. Sci.* 23, 1041–1057
50. Christoff, K. *et al.* (2016) Mind-wandering as spontaneous thought: A dynamic framework. *Nat. Rev. Neurosci.* 17, 718–731
51. Mattar, M.G. and Daw, N.D. (2018) Prioritized memory access explains planning and hippocampal replay. *Nat. Neurosci.* 21, 1609–1617
52. Kay, K. *et al.* (2020) Constant sub-second cycling between representations of possible futures in the hippocampus. *Cell* 180, 552-567.e25
53. Giallanza, T. *et al.* (2023) An integrated model of semantics and control*PsyArXiv*
54. Lundin, N.B. *et al.* (2023) Neural evidence of switch processes during semantic and phonetic foraging in human memory. *Proc. Natl. Acad. Sci. U. S. A.* 120

55. Fradkin, I. and Eldar, E. (2022) Accumulating evidence for myriad alternatives: Modeling the generation of free association. *Psychol. Rev.* DOI: 10.1037/rev0000397
56. Barrett, L.F. and Bar, M. (2009) See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1325–1334
57. Wood, W. and Neal, D.T. (2007) A new look at habits and the habit-goal interface. *Psychol. Rev.* 114, 843–863
58. Gross, J.J. (1999) Emotion and Emotion Regulation. In *Handbook of Personality: Theory and Research*, pp. 525–552, Guilford Publications
59. Lebreton, M. *et al.* (2009) An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron* 64, 431–439
60. Johnson, E.J. *et al.* (2007) Aspects of endowment: A query theory of value construction. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 461–474
61. Wang, S. *et al.* (2022) Mixing memory and desire: How memory reactivation supports deliberative decision-making. *Wiley Interdiscip. Rev. Cogn. Sci.* 13, e1581
62. Slovic, P. *et al.* (2007) The affect heuristic. *Eur. J. Oper. Res.* 177, 1333–1352
63. Liberman, N. and Trope, Y. (2014) Traversing psychological distance. *Trends Cogn. Sci.* 18, 364–369
64. Nook, E.C. *et al.* (2023) Emotion regulation is associated with increases in linguistic measures of both psychological distancing and abstractness*PsyArxiv 10.31219/osf.io/a2zv3*
65. Kross, E. and Ayduk, O. (2017) Self-Distancing. In *Advances in Experimental Social Psychology*, pp. 81–136, Elsevier
66. FitzGibbon, L. *et al.* (2020) The seductive lure of curiosity: information as a motivationally salient reward. *Curr. Opin. Behav. Sci.* 35, 21–27
67. Sharot, T. and Sunstein, C.R. (2020) How people decide what they want to know. *Nat. Hum. Behav.* 4, 14–19
68. Lewin, K. (1935) *A Dynamic Theory Of Personality*, Mcgraw-Hill
69. Mobbs, D. and Kim, J.J. (2015) Neuroethological studies of fear, anxiety, and risky decision-making in rodents and humans. *Curr. Opin. Behav. Sci.* 5, 8–15
70. McNaughton, N. *et al.* (2016) Approach/Avoidance. In *Neuroimaging Personality, Social Cognition, and Character*, pp. 25–49, Elsevier
71. Shell, D.F. (2023) Outcome expectancy in social cognitive theory: The role of contingency in agency and motivation in education. *Theory Pract.* 62, 255–265
72. Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215
73. Anderson, C.J. (2003) The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychol. Bull.* 129, 139–167
74. Heffner, J. *et al.* (2021) Emotion prediction errors guide socially adaptive behaviour. *Nat. Hum. Behav.* 5, 1391–1401

75. Corlazzoli, G. *et al.* (2023) Feeling and deciding: Subjective experiences rather than objective factors drive the decision to invest cognitive control. *Cognition* 240, 105587
76. Charpentier, C.J. *et al.* (2016) Models of affective decision making: How do feelings predict choice? *Psychol. Sci.* 27, 763–775
77. Fromer, 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
78. Dignath, D. *et al.* (2020) Conflict monitoring and the affective-signaling hypothesis-An integrative review. *Psychon. Bull. Rev.* 27, 193–216
79. Spunt, R.P. *et al.* (2012) The phenomenology of error processing: the dorsal ACC response to stop-signal errors tracks reports of negative affect. *J. Cogn. Neurosci.* 24, 1753–1765
80. Inzlicht, M. *et al.* (2015) Emotional foundations of cognitive control. *Trends Cogn. Sci.* 19, 126–132
81. Shenhav, A. *et al.* (2021) Decomposing the Motivation to Exert Mental Effort. *Curr. Dir. Psychol. Sci.* 30, 307–314
82. Inzlicht, M. *et al.* (2021) Integrating models of self-regulation. *Annu. Rev. Psychol.* 72, 319–345
83. Melnikoff, D.E. and Bargh, J.A. (2018) The mythical number two. *Trends Cogn. Sci.* 22, 280–293
84. Fudenberg, D. and Levine, D.K. (2006) A dual-self model of impulse control. *Am. Econ. Rev.* 96, 1449–1476
85. Alós-Ferrer, C. and Strack, F. (2014) From dual processes to multiple selves: Implications for economic behavior. *J. Econ. Psychol.* 41, 1–11
86. Schneider, I.K. and Schwarz, N. (2017) Mixed feelings: the case of ambivalence. *Curr. Opin. Behav. Sci.* 15, 39–45
87. Koob, G.F. and Le Moal, M. (1997) Drug abuse: hedonic homeostatic dysregulation. *Science* 278, 52–58
88. Levy, D.J. and Glimcher, P.W. (2012) The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* 22, 1027–1038
89. Padoa-Schioppa, C. (2011) Neurobiology of economic choice: a good-based model. *Annu. Rev. Neurosci.* 34, 333–359
90. Usher, M. and Mcclelland, J.L. (2001) The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* 108, 550–592
91. Cisek, P. (2012) Making decisions through a distributed consensus. *Curr. Opin. Neurobiol.* 22, 927–936
92. Hayden, B.Y. and Niv, Y. (04/2021) The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behav. Neurosci.* 135, 192–201
93. Walasek, L. and Brown, G.D.A. (2023) Incomparability and incommensurability in choice: No common currency of value? *Perspect. Psychol. Sci.* DOI: 10.1177/17456916231192828

94. Loomes, G. and Sugden, R. (1982) Regret theory: An alternative theory of rational choice under uncertainty. *Econ. J. Nepal* 92, 805–824
95. Sagi, A. and Friedland, N. (2007) The cost of richness: the effect of the size and diversity of decision sets on post-decision regret. *J. Pers. Soc. Psychol.* 93, 515–524
96. Peters, J. and Buchel, C. (2010) Neural representations of subjective reward value. *Behav. Brain Res.* 213, 135–141
97. Leng, X. *et al.* (2021) Dissociable influences of reward and punishment on adaptive cognitive control. *PLoS Comput. Biol.* 17, e1009737
98. Frömer, R. *et al.* (2022) Common neural choice signals emerge artifactually amidst multiple distinct value signals. BioRxiv 502393 [Preprint]. October 13, 2022. Available from: https://doi.org/10.1101/2022.08.02.502393
99. Zhang, Y. *et al.* (2023) Make or break: The influence of expected challenges and rewards on the motivation and experience associated with cognitive effort exertion*bioRxiv*
100. Saunders, B. *et al.* (2015) What does cognitive control feel like? Effective and ineffective cognitive control is associated with divergent phenomenology. *Psychophysiology* 52, 1205–1217
101. Quigley, K.S. *et al.* (2014) Inducing and measuring emotion and affect. In *Handbook of Research Methods in Social and Personality Psychology* (Reis, H. T. and Judd, C. M., eds), pp. 220–252, Cambridge University Press
102. FeldmanHall, O. and Heffner, J. (2022) A generalizable framework for assessing the role of emotion during choice. *Am. Psychol.* 77, 1017–1029
103. Knutson, B. *et al.* (2014) Inferring affect from fMRI data. *Trends Cogn. Sci.* 18, 422–428
104. Čeko, M. *et al.* (2022) Common and stimulus-type-specific brain representations of negative affect. *Nat. Neurosci.* 25, 760–770
105. Koban, L. *et al.* (2023) A neuromarker for drug and food craving distinguishes drug users from non-users. *Nat. Neurosci.* 26, 316–325
106. Abdel-Ghaffar, S.A. *et al.* (2024) Occipital-temporal cortical tuning to semantic and affective features of natural images predicts associated behavioral responses. *Nat. Commun.* 15, 5531
107. Kurzban, R. *et al.* (2012) A Cost/Benefit Model of Subjective Effort and Task Performance. *Behav. Brain Sci.* DOI: papers2://publication/uuid/90354C0F-9281-47F5-B68A-5E767AAD3D58
108. Hull, C.L. (1932) The goal-gradient hypothesis and maze learning. *Psychol. Rev.* 39, 25
109. Touré-Tillery, M. and Fishbach, A. (2011) The course of motivation. *J. Consum. Psychol.* 21, 414–423
110. Emanuel, A. *et al.* (2022) Why do people increase effort near a deadline? An opportunity-cost model of goal gradients. *J. Exp. Psychol. Gen.* 151, 2910–2926
111. Atkinson, J.W. (1957) Motivational determinants of risk-taking behavior. *Psychol. Rev.* 64, 359–372

112. Hockey, G.R.J. (2011) A motivational control theory of cognitive fatigue. In *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications. Decade of Behavior/Science Conference.* (Ackerman, P. L., ed), pp. 167–187, American Psychological Association
113. Carver, C.S. and Scheier, M.F. (1982) Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. *Psychol. Bull.* 92, 111–135
114. Solway, A. and Botvinick, M.M. (2012) Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychol. Rev.* 119, 120–154
115. Ritz, H. *et al.* (2022) Cognitive Control as a Multivariate Optimization Problem. *J. Cogn. Neurosci.* 34, 569–591
116. Juechems, K. and Summerfield, C. (2019) Where does value come from? *Trends Cogn. Sci.* 23, 836–850
117. Keramati, M. and Gutkin, B. (2014) Homeostatic reinforcement learning for integrating reward collection and physiological stability. *Elife* 3
118. Gunantara, N. (2018) A review of multi-objective optimization: Methods and its applications. *Cogent Eng.* 5, 1502242
119. Botvinick, M.M. *et al.* (2001) Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652
120. Hazy, T.E. *et al.* (2006) Banishing the homunculus: making working memory work. *Neuroscience* 139, 105–118
121. O'Reilly, R.C. (2020) Unraveling the mysteries of motivation. *Trends Cogn. Sci.* 24, 425–434
122. Kruglanski, A.W. *et al.* (2002) A theory of goal systems. In *Advances in Experimental Social Psychology*, pp. 331–378, Elsevier
123. Treadway, M.T. and Zald, D.H. (2011) Reconsidering anhedonia in depression: Lessons from translational neuroscience. *Neurosci. Biobehav. Rev.* 35, 537–555
124. Husain, M. and Roiser, J.P. (8/2018) Neuroscience of apathy and anhedonia: a transdiagnostic approach. *Nat. Rev. Neurosci.* 19, 470–484
125. Dalley, J.W. *et al.* (2011) Impulsivity, Compulsivity, and Top-Down Cognitive Control. *Neuron* 69, 680–694
126. Reyna, V.F. and Farley, F. (2006) Risk and Rationality in Adolescent Decision Making: Implications for Theory, Practice, and Public …. *Psychol. Sci. Public Interest* DOI: papers2://publication/uuid/E396F91F-E294-42AC-9AF4-929ED9E48483
127. Watson, D. *et al.* (1988) Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* DOI: papers://7E059DFB-CCD8-4D1E-964F-6D244D60A241/Paper/p201

128. Kouimtsidis, C. *et al.* (2014) How important are positive and negative outcome expectancies in the treatment of addiction: a narrative review of the literature. *Drugs Alcohol Today* 14, 137–149
129. Tsai, J.L. (2017) Ideal affect in daily life: implications for affective experience, health, and social behavior. *Curr. Opin. Psychol.* 17, 118–128
130. Tamir, M. (2009) What do people want to feel and why? *Curr. Dir. Psychol. Sci.* 18, 101–105