

# A normative perspective on motivation

Yael Niv<sup>1,3</sup>, Daphna Joel<sup>2</sup> and Peter Dayan<sup>3</sup>

<sup>1</sup> Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel

<sup>2</sup> Psychology Department, Tel Aviv University, Israel

<sup>3</sup> Gatsby Computational Neuroscience Unit, UCL, London, UK

**Understanding the effects of motivation on instrumental action selection, and specifically on its two main forms, goal-directed and habitual control, is fundamental to the study of decision making. Motivational states have been shown to ‘direct’ goal-directed behavior rather straightforwardly towards more valuable outcomes. However, how motivational states can influence outcome-insensitive habitual behavior is more mysterious. We adopt a normative perspective, assuming that animals seek to maximize the utilities they achieve, and viewing motivation as a mapping from outcomes to utilities. We suggest that habitual action selection can direct responding properly only in motivational states which pertained during behavioral training. However, in novel states, we propose that outcome-independent, global effects of the utilities can ‘energize’ habitual actions.**

## Introduction

Motivation occupies center stage in the psychology and behavioral neuroscience of decision making, and specifically instrumental action selection. There has been a recent renaissance in sophisticated analyses of motivation, primarily because manipulations such as specific satiety or motivational shifts have been used to tease apart different types of instrumental behaviors, namely, ‘goal-directed’ and ‘habitual’ control. These suggest that goal-directed and habitual actions are distinguished by the former’s, but not the latter’s, sensitivity to the utility of their specific outcomes [1]. Although goal-directed and habitual behavior can be characterized by their differing motivational sensitivities, and the effects of motivational manipulations on goal-directed behavior are relatively clear, exactly how (and indeed, whether) motivation influences habitual responding has remained unresolved. This is particularly disturbing as habitual responding plays a very prominent part in both normal and abnormal behavior.

That our understanding of motivational control is lacking might be partly because motivation itself is not a unitary construct [2]. In fact, Dickinson and Balleine [1] trace back to Descartes two very distinct influences of motivation on behavior: a ‘directing’ effect, determining the current goal(s) of behavior (e.g. food or water), and an ‘energizing’ effect, which determines the force or vigor underlying those actions. The latter is closely linked to Hullian ‘generalized drive’ [3–5], a motivational process that serves to energize all pre-potent actions. Whereas much is known about the directing aspects of motivation,

the ‘energizing’ effects of generalized drive have remained highly controversial.

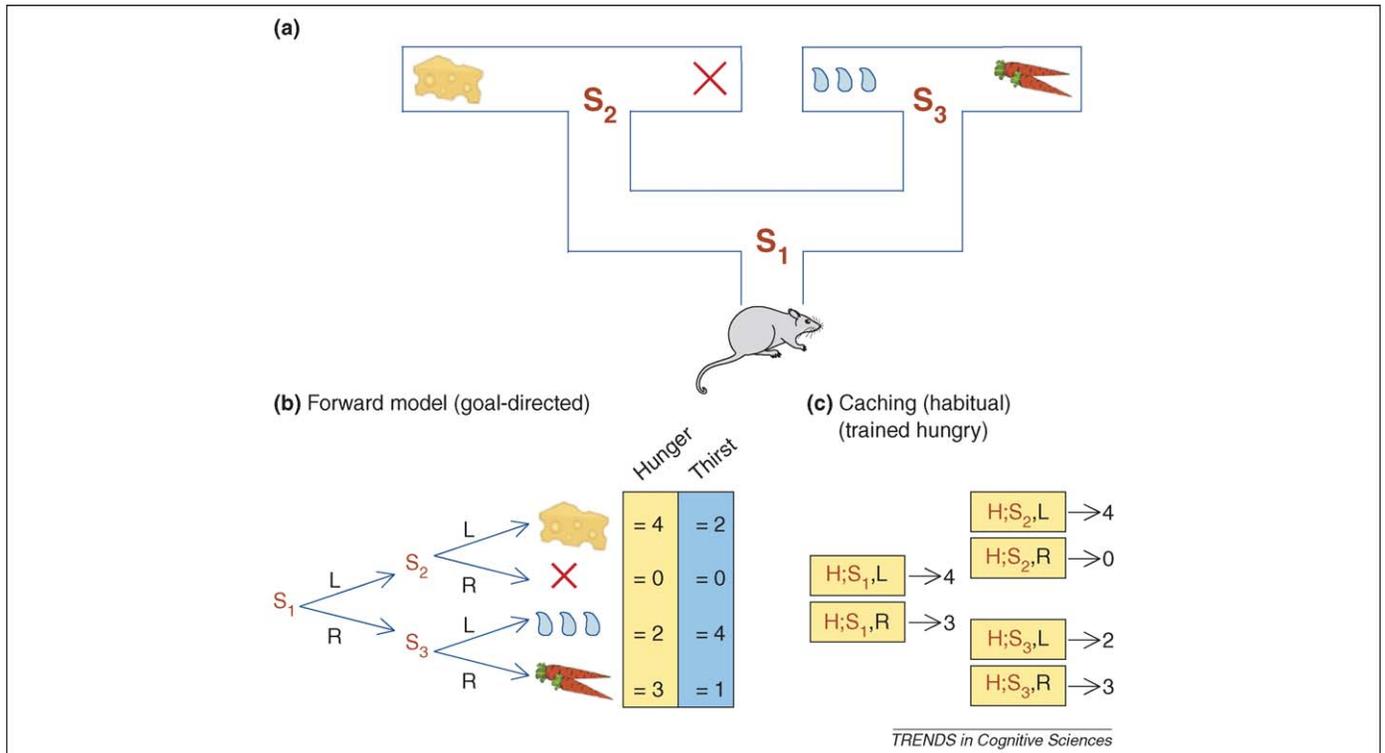
Here, we confront this challenge. We start by suggesting a simple, normative notion of motivation that allows us to define precisely outcome-specific ‘directing’ effects and outcome-independent ‘energizing’ effects. We then suggest that the outcome-specific effects of a novel motivational state predominantly influence goal-directed behavior, whereas the ‘energizing’ effects of generalized drive are seen in habitual responding [6]. As only preliminary experimental results on the latter hypothesis exist, we describe how it can best be tested, and detail its implications for both the understanding of motivational control and the resolution of the age-old debate regarding the existence of generalized drive.

## Motivation: a mapping from outcomes to utilities

Our conception of motivation is strongly influenced by the field of reinforcement learning [7]. In reinforcement learning, outcomes such as food or water have numerical utilities, and the imperative is to choose actions to maximize a long-term measure of total utility. However, in different motivational states, outcomes may have different utilities. We therefore define motivation as the *mapping* between outcomes and their utilities, and refer to ‘motivational states’ (e.g. ‘hunger’ or ‘thirst’) as indices of different such mappings (such as one in which foods are mapped to high utilities, and another in which liquids have high utilities). ‘Motivational shifts’ will refer to shifts between different motivational states. This is a pragmatic rather than philosophical definition, avoiding, for the moment, important issues such as the grounding of these mappings in evolutionary fitness. The definition is also means-neutral, in that organisms need not know these utilities, or have these utilities affect behavior in any way. Even if a dehydrated worm does not know the utility of different locations in terms of hydrating it, or how to get to those locations, by mere definition of ‘thirst’, some locations are now worth more than others. This sort of abstraction is useful, as it can be used to consider circumstances in which the different action selection systems do not have access to or knowledge of the true mapping, and can only approximate it.

How can an animal modify its behavior so as to maximize the utility it gains from its environment given its motivational state? This problem is especially challenging in tasks for which outcomes are dependent on whole sequences of action choices. Consider a hungry and a thirsty rat navigating a maze with food and water at

Corresponding author: Niv, Y. (yaelniv@alice.nc.huji.ac.il).



**Figure 1. Two strategies to solve the sequential action selection problem.** (a) A hypothetical problem: a rat navigates a maze with different outcomes at different end points. The rat starts at state  $S_1$  and must choose either left (L) or right (R). It must choose again at either  $S_2$  or  $S_3$ , to turn L or R to harvest one outcome. (b) By learning a forward model of the environment (essentially a state–action–outcome tree), the rat can decide whether to turn L or R at  $S_1$  by searching through the tree (simulating its next action choices) and finding the path with the highest overall utility. Crucially, the current motivational state of the rat defines the relevant mapping between outcomes and utilities (numbers in boxes), such that when hungry (yellow), the rat will find choice L optimal at  $S_1$ , but when thirsty (blue), it will prefer R. Behavior is thus goal-directed. (c) By contrast, a caching system does not represent the forward model, but rather stores (caches) learned values (in common-currency units) for every action at every state. After many learning trials, for each (state,action) pair, these approximate the experienced sum utilities of the outcomes which were eventually reached after taking this action at this state. Action selection simply involves choosing the action with the greatest cached value at the current state. Because the values are divorced from the identities of the outcomes produced by different actions, changes in the outcome–utility mapping cannot be translated into appropriate changes in values. However, the motivational state (hunger, H) can be stored as part of the state representation. In this way, action selection can be modified to match a different motivational mapping (e.g. relevant to thirst, T) if the set of (state,action) values relevant to that state  $\{(T;S_1,R),(T;S_2,L),\dots\}$  has previously been learned.

different locations (Figure 1a). Given the different utilities for the outcomes, how can the rats decide whether to turn left or right at the first choice point and how fast to run?

There is extensive evidence [1] that mice, rats and primates solve this problem using two neurally distinct [8] action selection schemes (in computational terms, two different controllers), which use different strategies [9,10]. The first, goal-directed action selection, driven by ‘response–outcome’ associations [1,11], is sensitive to the contingencies between actions and their outcomes, and to the utilities of these outcomes. The second, habitual action selection, is driven by ‘stimulus–response’ links [9], or, in computational terminology, stimulus–action values (or advantages) [12,13], and specifies actions without regard to their consequential outcomes. Box 1 discusses these two controllers in more detail, along with key findings about their inter-relationship and neural underpinnings. Below, we discuss how each action selection scheme can be influenced by motivation. We show that the division between outcome-specific ‘directing’ and general ‘energizing’ effects of motivation fits computationally and psychologically with the division between goal-directed and habitual controllers.

#### Goal-directed behavior: a ‘brute force’ solution

Almost by definition, the goal-directed system uses what is called a ‘forward model’, working out the ultimate

outcomes consequent on a sequence of actions by searching through the tree of state–actions–consequences, and choosing actions based on the outcomes’ current utilities (Figure 1b) [10]. Specific satiety and conditioned taste-aversion procedures (Box 2) have shown that action choice in this system is sensitive to manipulations that alter outcome utilities [14–21]. Further, studies introducing motivational shifts have shown that these too affect goal-directed behavior through the determination of outcome utilities. This is demonstrated by the fact that after a motivational shift, the new utilities must be experienced (in what is called an ‘incentive learning’ stage), for the effects of the motivational shift to be manifest [1,11,15,22–26].

Goal-directed control is therefore motivationally straightforward, with outcome utilities directing actions to the most valued outcomes appropriately. However, this form of search in a forward model constitutes a ‘brute-force’ solution to the action selection problem, involving high costs of computation and working memory, and is often intractable [10].

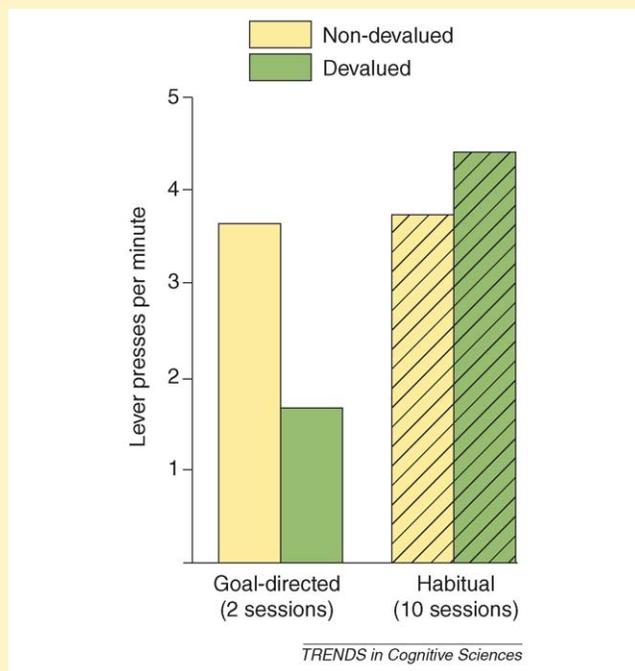
#### Is habitual behavior doomed to be motivation-insensitive?

Normative computational models of habitual action selection view it as arising from stored (cached) values of different actions in different states (Figure 1c). Each value

### Box 1. Goal-directed, and habitual action selection

By definition, goal-directed behavior is performed to obtain a desired goal. Although all instrumental behavior is **instrumental** in achieving its contingent goals, it is not necessarily purposively **goal-directed**. Dickinson and Balleine [1,11] proposed that behavior is goal-directed if: (i) it is sensitive to the contingency between action and outcome, and (ii) the outcome is desired. Based on the second condition, motivational manipulations have been used to distinguish between two systems of action control: if an instrumental outcome is no longer a valued goal (for instance, food for a satiated animal) and the behavior persists, it must *not* be goal-directed. Indeed, after moderate amounts of training, outcome revaluation (Box 2) brings about an appropriate change in instrumental actions (e.g. leverpressing) [43,44], but this is no longer the case for extensively trained responses ([30–31], but see [45]). That extensive training can render an instrumental action independent of the value of its consequent outcome has been regarded as the experimental parallel of the folk psychology maxim that well-performed actions become **habitual** [9] (see Figure 1).

This distinction between two types of behavior is also paralleled by a distinction between two different neural pathways to action selection. Habitual behavior is thought to be dependent on the dorsolateral striatum [8,32] and its dopaminergic afferents, whereas goal-directed behavior is controlled more by circuitry involving frontal cortical areas and the dorsomedial striatum [8,20,21]. These two pathways have been suggested as subserving two action controllers with different computational characteristics, which operate in parallel during action selection [10].



**Figure 1. Habitual and goal-directed behaviors.** When hungry rats are trained to press a lever to obtain sucrose pellets, post-training devaluation of the pellets by conditioning taste aversion (green bars) causes a reduction in lever-pressing, compared with rats for whom the outcome was not devalued (yellow bars), but only after moderate training when responding is still goal-directed (left). After considerable training, the behavior becomes habitual (right, hatched) and insensitive to the utility of the outcome. In all cases behavior was tested in extinction (i.e. with no pellets provided). Adapted from [9] with permission.

is defined in terms of the expected cumulative future utilities consequent on performing this action in this state. Adding together the utilities of different outcomes (food, drink, mates, etc.), cached values are thus outcome-general and defined in units of a common currency. The values are acquired through extensive experience by a process of model-free reinforcement learning [7,10], using methods

such as temporal difference learning [10,27,28]. To deal with potentially long sequences of actions, these methods care only about accumulated utilities. Specifically, they avoid building a forward model such as that in Figure 1b, and pay no regard to the *identity* of the actual outcomes consequent on the actions chosen. At decision points, actions are chosen by comparing their relative cached values, rather than their consequent outcomes. Although less powerful than methods involving forward models, this sort of action control offers substantial computational savings. This underlies its popularity in reinforcement learning. Further, the neurobiological substrate of such methods has been intensively investigated [28,29].

By contrast with goal-directed actions, habitual behavior is operationally defined by its very insensitivity to its consequent outcomes. How, then, can habits be influenced by a change in the motivational mapping of outcomes to utilities? One straightforward way is through learning of new values, based on experiencing the new utilities. Moreover, the motivational state effective at the time of learning can be used to index values learned in different states, and keep them separate (Figure 1c). In this way, habitual behavior can indirectly learn a motivation-dependent behavioral policy which properly directs actions to maximize outcome utility in different (but known) motivational states.

But what about the immediate effects of new outcome utilities which have never been experienced in the task? The question of how untrained outcome utilities affect habits touches directly upon the core issue of motivational control of habits (and necessitates the use of extinction tests; Box 2). Unfortunately, the literature is divided on this – some studies show insensitivity to outcome devaluations [17,18,30–32]; whereas others claim that habitual behavior is *directly* sensitive to motivational manipulations [33]. We suggest that this confusion stems from treating outcome revaluation by a motivational shift as equivalent to outcome revaluation by specific satiety or conditioned taste-aversion (Box 2). Indeed, unlike goal-directed control, habitual control cannot direct action selection according to new outcome utilities without the learning of new values described above, explaining the lack of sensitivity to the latter outcome devaluation procedures. Nevertheless, in the case of motivational shifts, we claim that even without new learning, habitual behavior can be partially adapted using two different well-founded approximations to the desired effects of the new outcome-utility mapping. One involves a form of generalization gradient, based on an internal representation of the motivational state; the other involves a form of immediate ‘generalized drive’ effect on ongoing behavior [6]. We describe these approximations in turn.

#### Approximation 1: Generalization decrement

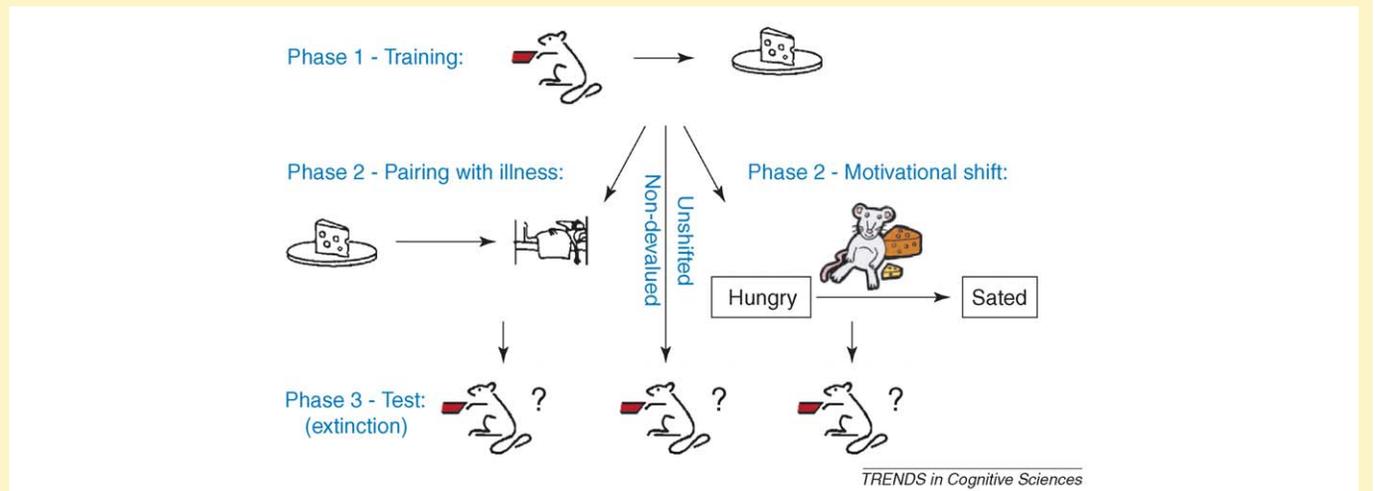
We argued above that the cached values can be indexed by the current motivational state (Figure 1c). In this case, any change in motivational state from training to test will potentially lead to a change in the estimated action values. Given the evidence for generalization decrement following a change in stimuli between training and test [34] (i.e. a reduction in responding when tested with stimuli different

## Box 2. Methods for outcome revaluation

Post-training reinforcer revaluations have proved invaluable for the study of effects of motivation on action selection [1,46]. In a typical experiment (Figure 1), food-deprived rats are trained to perform an instrumental action (such as leverpressing) to obtain a rewarding outcome (food). After behavior has been acquired, a post-training stage modifies the value of the outcome for one group of rats. The consequences of this manipulation are tested by comparing the propensity of these rats to perform the instrumental response, with that of rats for whom the outcome has not been revalued. Importantly, this is done in extinction, that is, in the absence of rewards, to test for the effects of the revaluation on the previously learned associations, and avoid new learning. A significant difference in responding is evidence for sensitivity to the change in the value of the outcome.

Three methods are commonly used for outcome revaluation: In a **specific satiety procedure** [15–18,20,21,47], the rats are pre-fed on the outcome, such that they develop a temporary, outcome specific

satiation for this outcome. Consumption tests show that such a procedure selectively devalues only the pre-fed outcome. Another method for devaluing a specific outcome is by **conditioning taste aversion** to it [19,31,32,43,45,47]. In this procedure, after the rat consumes the outcome, gastric illness is induced, rendering the food aversive to the rat. Finally, **motivational shifts** [15,22–25,33] can either devalue or enhance the value of outcomes. Most commonly, after training rats to leverpress when hungry, their motivational state is shifted to that of satiety by allowing consumption *ad lib* of laboratory-chow in the home-cage. This manipulation renders the once very valuable food reward less valuable. Opposite shifts (from training when sated to testing when hungry) enhance the value of the instrumental outcome, and shifts between different motivational states (for instance, between hunger and thirst) can be used to change the value of one outcome (say, food pellets) while maintaining the value of another (e.g. sucrose solution).



**Figure 1. Experimental techniques for outcome revaluation.** In a typical outcome revaluation experiment, rats are first trained (Phase 1) to perform an instrumental action (here, pressing a lever) to obtain a desired outcome. In Phase 2 the outcome value is manipulated by, for example, pairing its consumption with illness (left) or inducing a motivational shift, such as from hunger to satiety (right). In Phase 3 the trained response is tested in extinction (i.e. with no outcomes available), and behavior of rats for which the outcome has been revalued is compared with that of rats who have not undergone Phase 2. Rat cartoons courtesy of Bernard Balleine.

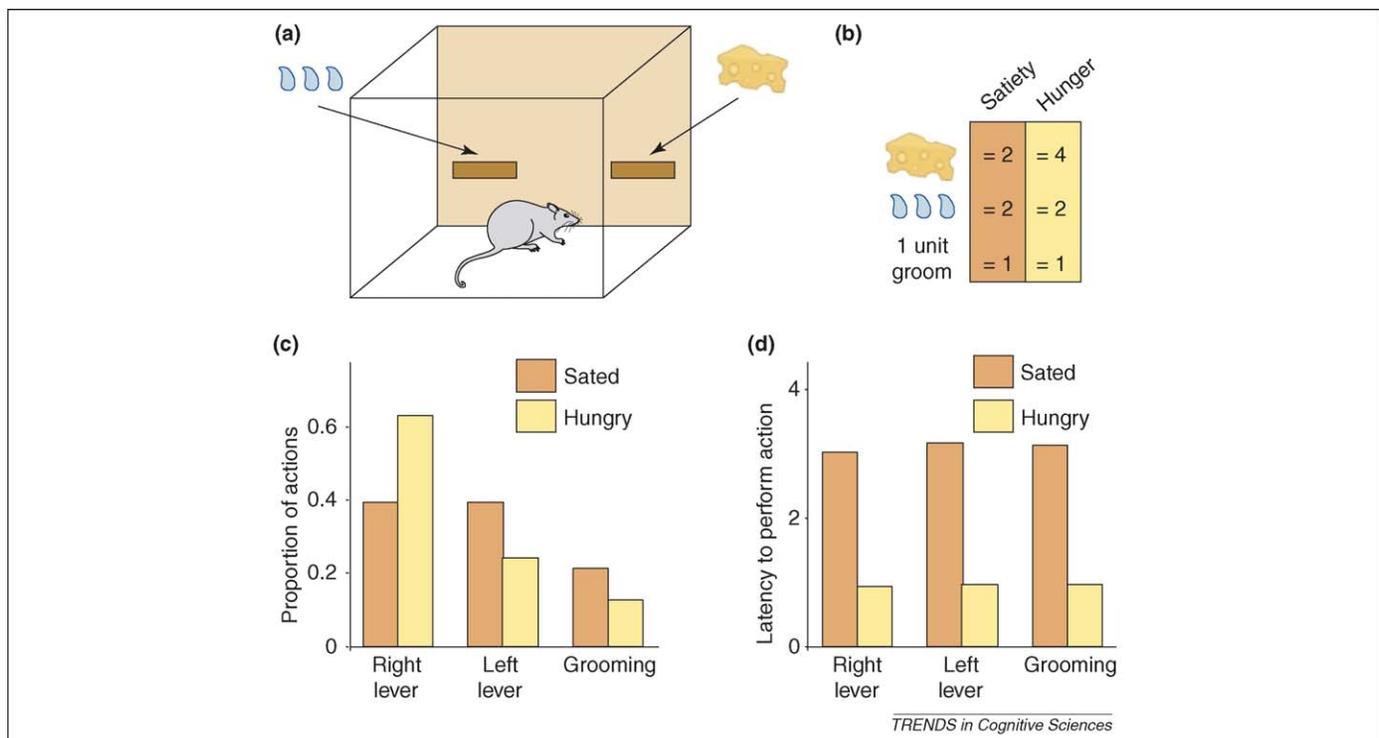
from those with which the behavior was trained), one may expect that a shift to a novel motivational state may also lead to decreases in responding [1,4,5].

### Approximation 2: Generalized drive

The second form of generalization stems from the fact that outcomes tend to have higher utilities in more deprived states, making the expected average reward per unit time higher. According to a recent normative model of free operant behavior [6], this average reward rate plays an important role in determining optimal response rates. In the model, the optimal rates of performing actions (Figure 2a) are calculated based on the utilities of the outcomes, and the assumed costs of acting quickly. It turns out that enhancing the utility of a subset of outcomes (say, food, as a result of hunger, Figure 2b), has two different consequences. First, actions leading to these outcomes are chosen more often (Figure 2c), as in the directing effect of motivation. Second, *all* actions are performed at a faster rate regardless of the identity of their outcome (Figure 2d). This happens because the average reward rate constitutes a form of ‘opportunity cost’ on response latencies, defining how much reward is forfeited in every idle second. So when the average reward rate is higher, the higher cost of sloth

induces more overall rapid responding. This ‘generalized drive’ effect of higher deprivation can be seen as being orthogonal to the directing effect (different from the suggestion that incentive motivation to a specific outcome energizes actions leading to it [1]). This is because in the model the choice *between* actions is only affected by their specific outcomes, whereas the choice of *how fast* to perform the selected action is dependent only on the average reward rate.

In motivational states such as hunger or thirst, in which the average reward rate is high (because the utilities of food or fluid outcomes are high), the model predicts that all pre-potent actions should be performed faster. In states such as satiety, with lower average rates of reward, all pre-potent actions should be performed more slowly. Therefore, provided only that it has an idea as to whether the average rate of reward in a new motivational state will be higher or lower, the habitual system can respond approximately appropriately, by modulating the rate of performance of all actions regardless of their consequences. This result gives the old (and controversial) psychological notion of ‘generalized drive’ [3–5] a new, normative interpretation, as an optimal solution to an action-selection problem. By incorporating sensitivity to average reward rates in



**Figure 2. Two behavioral consequences of a motivational shift.** (a) A simulated rat, trained in an operant chamber, has three choices: pressing the left lever to obtain water, pressing the right lever to obtain cheese, or grooming to obtain some internal reward. (b) Even when relatively sated, the cheese and water have slightly higher utilities than grooming. A shift to hunger, however, markedly enhances the utility of cheese compared with the other utilities. (c) Unsurprisingly, as a result of the shift from satiety to hunger, the rat chooses to press the right lever (to obtain cheese) more often than either grooming or pressing the left lever (which are still performed, albeit less often). (d) A second consequence of the motivational shift is that all actions are now performed faster (latency in arbitrary units). Hence, grooming and pressing the left lever, when performed, are also executed faster. This ‘energizing’ effect of the motivational shift is thus not specific to the action leading to the favored outcome, and can be regarded an outcome-independent effect. Data in (c) and (d) redrawn with permission from [6].

determining rates of responding, the habitual system can immediately at least approximate the optimal choices of response *rates*, even if not the actual optimal actions. Of course, given additional training, this approximation will be refined and action selection will become precisely correct once the new values are learned.

This notion of a generalized drive effect of motivational shifts explains the observation that habitual responding is directly sensitive to motivational shifts [33]. It is also not surprising that this does not necessitate an ‘incentive learning’ stage, as the effect is presumably not modulated by a specific change in outcome utility. However, that particular study [33] did not examine leverpressing for an outcome whose utility was left fixed by the motivational shift (e.g. water), which would prove the real test case for the form of ‘generalized drive’ hypothesis that we are suggesting.

In summary, there are at least three possible reasons for a reduction in habitual responding after a shift from hunger to satiety – a generalized drive effect, an outcome-specific effect (i.e. a decrease in the outcome’s utility), and generalization decrement. Box 3 details how the use of motivational upshifts and side-shifts, as well as training with several different outcomes, can tease these effects apart, and make a conclusive case for or against our generalized drive hypothesis. Preliminary results from our laboratory (Y. Niv *et al.*, unpublished report: <http://leibniz.cs.huji.ac.il/tr/857.pdf>) indeed support a role for both generalized drive and generalization decrement in habitual responding, and show no evidence for outcome-

specific effects for either motivational side-shifts or up-shifts.

### Two sides of motivational influence: the directing and the energizing

In summary, a normative analysis of the different revaluation manipulations used to establish the characteristics of habitual and goal-directed behavior suggests that the outcome-specific ‘directing’ effects of a novel motivational state influence goal-directed behavior, whereas the ‘energizing’ effects of generalized drive are seen in habitual responding. This distinction also calls for the operational definition of habitual behavior to be slightly refined. Habits are not in general insensitive to outcome revaluations, but only do not show outcome-specific sensitivity to such manipulations. Of course, theoretically, goal-directed behavior should also show outcome-independent energizing effects. However, as these might be overwhelmed by directing effects, teasing them apart will require a careful analysis of inter-response latencies (Box 3).

This division into outcome-dependent and outcome-independent effects has an interesting parallel in the phenomenon of Pavlovian-instrumental transfer (PIT). In PIT, stimuli classically conditioned to predict the occurrence of affectively significant outcomes affect the vigor of instrumental responding. As with motivational influences, there are two sorts of PIT: specific, in which a stimulus only affects instrumental responding for a similar outcome, and general, in which a stimulus has a general influence on all instrumental actions regardless of their outcome [19]. This

### Box 3. Teasing apart the effects of motivational shifts

Theoretically, there are three routes by which shifts in motivational state can modulate behavior. One is through modulation of the utility (or incentive value) of the goals of behavior [2,48,49]. This outcome-specific effect would be manifest in the 'directing' aspect of motivational control. A second route was proposed by Hull in his Generalized Drive hypothesis [3–5] in terms of the 'energizing' aspect of motivation. According to this, motivational states exert a certain 'drive' which is applicable to ongoing behavior. For instance, sated rats may be less inclined to perform any pre-potent action as a result of reduced generalized drive. Importantly, this effect is not outcome-specific [50]. Last, post-training shifts to a novel motivational state can influence behavior because of a generalization decrement from the training context (which potentially includes the motivational state) to the test context [1,4]. This effect is not only outcome-independent, but is also independent of the identity of the motivational state, and predicts a reduction in responding for any motivational shift (even an up-shift from low to high deprivation) [4].

These potential effects are not at all mutually exclusive; however, they can predict different directions of change of behavior, owing to their different dependencies on the identity of outcomes and

motivational states. Table I illustrates predictions for qualitatively different motivational shifts: a down-shift from a deprived to an undeprived state (e.g. from hunger to satiety), an up-shift (e.g. from satiety to thirst), and a side-shift between two different motivational states (e.g. from hunger to thirst). Predictions are illustrated for behavior whose outcome is either sucrose pellets (relevant only in hunger) or sucrose solution (relevant both in hunger and thirst). Arrows illustrate a predicted reduction, increase, or no change in rate of behavior as compared with unshifted controls. The prediction regarding the drive effect for side-shifts is undetermined, as it is not possible to measure independently the relative drive induced by hunger versus thirst.

By comparing the effects of different shifts on responding for two different outcomes as illustrated, the different contributions to motivational control of habits can be distinguished. Furthermore, to tap 'energizing' effects unconfounded by directing effects (which is especially important in goal-directed behavior where the latter are prominent), it is important to use behavioral measures such as inter-response latencies and not only overall response counts.

Table I. Predictions for the effects of motivational shifts

	Down-shift (hunger → satiety)	Up-shift (satiety → thirst)	Side-shift (hunger → thirst)
Outcome specific (‘directing’ effect)	↓	↔ or ↑ pellets solution	↓ or ↔ pellets solution
Drive (‘energizing’ effect)	↓	↑	?
Generalization decrement	↓	↓	↓

latter effect is reminiscent of the generalized drive effect which we have tied to average reward rates.

Building on what is known about the neural substrates of the two forms of PIT [35–38], as well as the substrates of goal-directed and habitual control (Box 1), we can now speculate as to the neural basis of the two forms of motivational influence. In accord with computational models [6,28], and the role of dopamine in habitual learning and action selection [39,40], we propose that the influence of generalized drive, or 'energizing' motivational effects on responding is dopamine-dependent [6,41], possibly mediated by the nucleus accumbens and the central nucleus of the amygdala [42]. By contrast, we speculate that 'directing' motivational control through determination of specific outcome values is probably dopamine independent, and possibly mediated by the posterior basolateral amygdala [42,8], and the orbitofrontal cortex [42]. Moreover, the suggested dopamine-dependence of generalized drive effects, tied with the demonstrated

dopamine-dependence of general PIT [35,36], prompts the tantalizing thought that the bases for the two may be the same, providing a potentially strong link between motivation and classical (Pavlovian) conditioning in controlling instrumental behavior (Box 4).

### Conclusions

Motivation turns out to be a rich and complex topic, because it has multiple facets to which the various action-selection systems are differentially sensitive. Oddly, it has been easier to use motivation to dissociate these systems than it has been to use them to elucidate motivation. Our definition of motivational states in terms of mappings between outcomes and utilities provides a simple normative scaffold on which to understand both optimal and approximately optimal sensitivity to outcome utilities. These ideas regarding the ways motivation influences action selection, and specifically habitual control, are not only significant for the understanding of motivation, but also provide a possible normative foundation for the much debated concept of generalized drive. The use of computational models grounds this concept in precise predictions about what the effects of generalized drive should be, and how they should be measured to tease them apart from qualitatively different, orthogonal effects of other aspects of motivation.

### Acknowledgements

We are grateful to Misha Ahrens, Bernard Balleine, Nathaniel Daw, Máté Lengyel, Ken Norman, Tom Schonberg, Ina Weiner and Louise Whiteley for helpful comments on earlier versions of the manuscript, and to Nathaniel Daw for much discussion and sharing of ideas. Our gratitude

### Box 4. Questions for future research

- What are the effects on habitual behavior of up-shifts and side-shifts to an untrained motivational state?
- Can 'generalized drive' effects be seen when measuring individual response latencies in goal-directed behavior?
- Is there a dissociation in terms of response controllers and motivational effects in Pavlovian behavior, similar to that in instrumental control?
- Do generalized drive effects and general Pavlovian-instrumental-transfer share a common neural substrate?
- Are 'directing' motivational effects and outcome-specific Pavlovian-instrumental-transfer mediated by similar neural mechanisms?

goes to Sharon Riwkes and Eran Katz who carried out some of the experiments on motivational control of habitual behavior. This research was funded by a Dan David fellowship and a Hebrew University Rector Fellowship to Y.N., and the Gatsby Charitable Foundation.

## References

- 1 Dickinson, A. and Balleine, B. (2002) The role of learning in the operation of motivational systems. In *Learning, Motivation and Emotion* (Vol. 3) (Gallistel, C.R., ed.), pp. 497–533, Wiley
- 2 Berridge, K.C. (2004) Motivation concepts in behavioral neuroscience. *Physiol. Behav.* 81, 179–209
- 3 Hull, C. (1943) *Principles of Behavior*, Appleton
- 4 Brown, J. (1961) *The Motivation of Behavior*, McGraw-Hill
- 5 Bolles, R. (1967) *Theory of Motivation*, Harper & Row
- 6 Niv, Y. et al. (2005) How fast to work: Response vigor, motivation and tonic dopamine. In *Advances in Neural Information Processing Systems* (Vol. 18) (Weiss, Y. et al., eds), pp. 1019–1026, MIT Press
- 7 Sutton, R.S. and Barto, A.G. (1998) *Reinforcement Learning*, MIT Press
- 8 Balleine, B.W. (2005) Neural bases of food seeking: Affect, arousal and reward in corticostriatal limbic circuits. *Physiol. Behav.* 86, 717–730
- 9 Dickinson, A. (1985) Actions and habits: The development of behavioral autonomy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 308, 67–78
- 10 Daw, N.D. et al. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711
- 11 Dickinson, A. and Balleine, B.W. (1994) Motivational control of goal-directed action. *Anim. Learn. Behav.* 22, 1–18
- 12 Watkins, C.J.C.H. and Dayan, P. (1992) Q-learning. *Mach. Learn.* 8, 279–292
- 13 O'Doherty, J. et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454
- 14 Balleine, B.W. and Dickinson, A. (1998) Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419
- 15 Balleine, B.W. and Dickinson, A. (2000) The effect of lesions of the insular cortex on instrumental conditioning: Evidence for a role in incentive memory. *J. Neurosci.* 20, 8954–8964
- 16 Corbit, L.H. et al. (2001) The role of nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *J. Neurosci.* 21, 3251–3260
- 17 Killcross, S. and Coutureau, E. (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* 13, 400–408
- 18 Coutureau, E. and Killcross, S. (2003) Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav. Brain Res.* 146, 167–174
- 19 Holland, P. (2004) Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *J. Exp. Psychol. Anim. Behav. Process.* 30, 104–117
- 20 Yin, H.H. et al. (2005a) Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.* 22, 505–512
- 21 Yin, H.H. et al. (2005b) The role of dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523
- 22 Dickinson, A. and Dawson, G. (1989) Incentive learning and the motivational control of instrumental performance. *Q. J. Exp. Psychol.* 41B, 99–112
- 23 Balleine, B.W. (1992) Instrumental performance following a shift in primary motivation depends on incentive learning. *J. Exp. Psychol. Anim. Behav. Process.* 18, 236–250
- 24 Lopez, M. et al. (1992) Incentive learning and the motivational control of instrumental performance by thirst. *Anim. Learn. Behav.* 20, 322–328
- 25 Balleine, B.W. et al. (1995) Motivational control of heterogeneous instrumental chains. *J. Exp. Psychol. Anim. Behav. Process.* 21, 203–217
- 26 Balleine, B.W. (2000) Incentive processes in instrumental conditioning. In *Handbook of Contemporary Learning Theories* (Mowrer, R. and Klein, S., eds), pp. 307–366, Erlbaum
- 27 Barto, A.G. et al. (1983) Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. Sys. Man and Cyber* 13, 834–846
- 28 Montague, P.R. et al. (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947
- 29 Schultz, W. et al. (1997) A neural substrate of prediction and reward. *Science* 275, 1593–1599
- 30 Adams, C. (1980) Post conditioning devaluation of an instrumental reinforcer has no effects on extinction performance. *Q. J. Exp. Psychol.* 32B, 447–458
- 31 Adams, C. (1982) Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol.* 34B, 77–98
- 32 Yin, H.H. et al. (2004) Lesions of the dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189
- 33 Dickinson, A. et al. (1995) Motivational control after extended instrumental training. *Anim. Learn. Behav.* 23, 197–206
- 34 Domjan, M. (2003) *Principles of Learning and Behavior*, (5th edn), Thomson/Wadsworth
- 35 Dickinson, A. et al. (2000) Dissociation of Pavlovian and instrumental incentive learning under dopamine agonists. *Behav. Neurosci.* 114, 468–483
- 36 Wyvell, C.L. and Berridge, K.C. (2000) Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward 'wanting' without enhanced 'liking' or response reinforcement. *J. Neurosci.* 20, 8122–8130
- 37 Holland, P.C. and Gallagher, M. (2003) Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned-stimulus potentiated feeding and Pavlovian-instrumental transfer. *Eur. J. Neurosci.* 17, 1680–1694
- 38 Corbit, L.H. and Balleine, B.W. (2005) Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer. *J. Neurosci.* 25, 962–970
- 39 Faure, A. et al. (2005) Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *J. Neurosci.* 25, 2771–2780
- 40 Salamone, J.D. and Correa, M. (2002) Motivational views of reinforcement: Implications for understanding the behavioral function of nucleus accumbens dopamine. *Behav. Brain Res.* 137, 3–25
- 41 Weiner, I. and Joel, D. (2002) Dopamine in schizophrenia: Dysfunctional information processing in basal ganglia-thalamocortical split circuits. In *Handbook of Experimental Pharmacology* (Vol. 154/II): *Dopamine in the CNS* (Chiara, G.D., ed.), pp. 417–472, Springer-Verlag
- 42 Cardinal, R.N. et al. (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26, 321–352
- 43 Adams, C. and Dickinson, A. (1981) Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol.* 33B, 109–121
- 44 Dickinson, A. and Nicholas, D. (1983) Irrelevant incentive learning during training on ratio and interval schedules. *Q. J. Exp. Psychol.* 35B, 235–247
- 45 Colwill, R. and Rescorla, R. (1988) The role of response-reinforcement associations increases throughout extended instrumental training. *Anim. Learn. Behav.* 16, 105–111
- 46 Dayan, P. and Balleine, B.W. (2002) Reward, motivation and reinforcement learning. *Neuron* 36, 285–298
- 47 Colwill, R. and Rescorla, R. (1985) Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *J. Exp. Psychol. Anim. Behav. Process.* 11, 520–536
- 48 Tolman, E.C. (1949) The nature and functioning of wants. *Psychol. Rev.* 56, 357–369
- 49 Bindra, D. (1974) A motivational view of learning, performance, and behavior modification. *Psychol. Rev.* 81, 199–213
- 50 Mackintosh, N.J. (1974) *The Psychology of Animal Learning*, Academic Press