The Rate-Correlation Theory of Goal-Directed Behavior: An Update

Omar D. Perez

Abstract Over half a century ago, Baum (J Exp Anal Behav 20:137–153, 1973) proposed a theory based on the idea that behavior should be analyzed in terms of rates of activities rather than single, discrete units of analysis. The evidence from animal experiments suggests that his theory could be underlying goal-directed control, in that actions are performed more often when there is a direct correlation between behavior rate and reinforcement rate in a given period of time and are more sensitive to changes in outcome value than when the rate correlation is weak. This correlational system, coupled with a reinforcement learning algorithm of habit learning, can capture a wide range of data from animal experiments. In this chapter, I discuss this theory in light of recent human data and possible extensions of this idea to other areas of research.

Keywords Goal-directed · Habit · Model-based · Model-free · Reinforcement learning

A paper published by Baum five decades ago presented an argument that called for a reassessment of the classic behaviorist idea which assigned a critical role for stimulus-response (S-R) learning to behavior. Baum's argument was based on compelling data indicating that another variable, the correlation between the rate of behavior and the rate of reinforcement, offered a more effective explanation of instrumental conditioning results across a wide range of studies and species (Baum, 1973). This theory challenged the prevailing notion that a subject's experience could be reduced to simple S-R links formed through contiguous reinforcement of specific behaviors in specific contexts. Instead, Baum proposed a more comprehensive perspective, arguing that behavior should be viewed as an activity that extends over time—the response rate—where multiple responses and reinforcers within a given period in a particular context contribute to strengthening the S-R link. Consequently,

O. D. Perez (✉)
Complex Engineering Systems Institute, Santiago, Chile

Department of Industrial Engineering, University of Chile, Santiago, Chile
e-mail: omar.perez.r@uchile.cl

Y. Vandaele (ed.), *Habits*, https://doi.org/10.1007/978-3-031-55889-4_4

behavior should not be considered as a collection of discrete events, but rather an ongoing process where the reinforcement rate serves as a nondiscrete reinforcer following the response rate in a feedback system between the organism and the environment (Baum, 2012).

1 Correlational Theory as Goal-Directed Behavior

Dickinson (1985) proposed a departure from Baum's correlational perspective, offering a novel approach without compromising the computational concept or explanatory potential of the idea. Instead, Dickinson argued that correlational theory would be best understood as the foundation for goal-directed control, distinct from the simplistic S-R reinforcement commonly referred to as a habit. In this context, an agent guided by correlations between events should encode the causal relationship between actions and impending outcomes by computing the correlation between response and reinforcer rates, while also incorporating the sensory properties of the obtained reinforcers. This forward-looking system should therefore incorporate the strength of the causal link as the rate correlation experienced by the subject and an incentive system that assigns utility or value for the outcomes produced by behavior. As a consequence, behavior should be sensitive to changes in the causal contingency between responses and reinforcers (the rate correlation) and also to changes in outcome value.

The primary method for assessing goal-directed control of behavior is through the implementation of an outcome devaluation procedure (Adams & Dickinson, 1981). This procedure is founded on the notion that in goal-directed behavior, alterations in the value of the outcome resulting from an action should directly and immediately impact performance, without requiring reexperiencing of the newly assigned value of the outcome upon performing the response. In a seminal study, Adams and Dickinson (1981) trained hungry rats to press a lever to obtain a rewarding outcome while a noncontingent alternative reinforcer was provided. To reduce the relative value of one outcome with respect to the other, they established a flavor aversion by associating its consumption with gastric malaise in a distinct context until the animals ceased consuming the reinforcer when freely presented (i.e., the devaluation manipulation reduced the value of the outcome). During the subsequent test phase, they allowed the animals to press the same lever as in training, but without delivering the outcome (i.e., extinction phase). Critically, they observed that the animals whose devalued outcome was contingent upon responding exhibited reduced lever-pressing compared to animals in which the noncontingent outcome was devalued. As the test was conducted during extinction, the change in behavior reflects knowledge acquired during the training phase rather than during the test itself. Clearly, the rats in the study considered the consequences of their actions and the rewarding value associated with the outcome.

Later on, Dickinson and his colleagues demonstrated that the response-outcome schedule in effect (ratio or interval schedule) could render behavior habitual or goal-

directed (Dickinson et al., 1983). On a ratio schedule of reinforcement, there is a fixed probability of reinforcement per action performed so that more effort (that is, a higher response rate) leads to a higher reinforcement rate. By contrast, under interval schedules the reinforcer becomes available on average after a period of time since the last delivered reinforcer, so that higher effort does not necessarily lead to higher rates of reinforcement. These studies showed that two behavioral systems could be engaged under different circumstances. This groundbreaking finding—which has received more recent replications (Gremel & Costa, 2013)—is the basis for a whole line of research in humans and animals to this day.

The first observation made by Dickinson et al. (1983) was that ratio schedules supported higher response rates than interval schedules when reward probabilities or rates were matched between groups. An explanation for this phenomenon was on the basis of the experienced response-reinforcer rate correlation, which should be positive for ratio schedules and weaker for interval schedules once the response rate reaches a level such that all reinforcers are collected as soon as they become available. More importantly, however, is the second observation from this study: ratio schedules of reinforcement were able to maintain goal-directed control more effectively than interval schedules. It appeared that the experienced rate correlation of ratio schedules was not only underlying behavioral performance but was also involved in goal-directed control.

Perez and Dickinson (2020) formalized this and other ideas in a theory including two concurrent systems that control behavior and collectively determine the observed level of responding across different experimental conditions. The first system, the goal-directed, is sensitive to the correlation between response rate and reinforcement rate, while the second system, the habitual, aligns with the established principles of reward prediction error which are present in model-free reinforcement learning algorithms (Bush & Mosteller, 1951; Sutton & Barto, 2018). In this habit system, it is the response-reinforcer contiguity—instantiated by reinforcer probability—that drives behavior, and the prediction from both systems determines the reward prediction error. Unlike the goal-directed system where events are considered as rates in time, in the habit system it is the relationship between each response and its reinforcement that drives changes in habit strength. If events are received with some delay, for example, the habit strength will be affected because there is no immediate reinforcement for those responses (a negative reward prediction error), but the correlational goal-directed system will still consider those events as being a consequence of the response rate in a given period of time. For the goal-directed system, what matters is the number of responses and reinforcers in time, not whether each of the responses was followed (or not) by a reinforcer.

Figure 1 presents a schematic representation of this theory in a dual-system model. In each memory recycle, the agent deploys a local memory of time samples (size $m = 5$ in the figure) to compute the correlation between events. On the next recycle, the agent randomly forgets one of the time samples and adds a new sample, from which a new local correlation is computed. Goal-directed strength is assumed to be a direct function of the experienced local rate correlation in each memory recycle. The habit system, by contrast, computes a prediction error at the level of

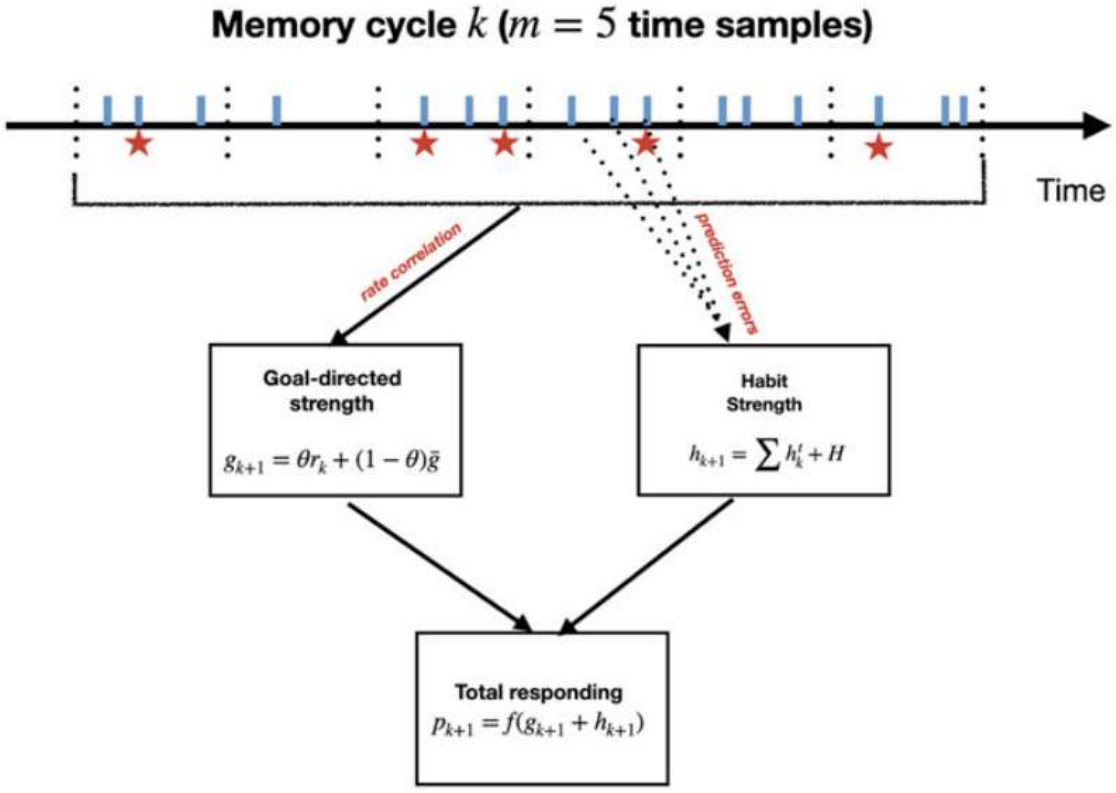

Fig. 1 Schematic representation of Perez and Dickinson's (2020) model. Agents deploy a mnemonic system of m time samples ($m = 5$ in this example) and experience a correlation between response and reinforcer rates (r_k) within each memory recycle k. On the next recycle $k + 1$, one of the time samples is randomly forgotten and a new experienced correlation is computed. This gives the agent a response strength from the goal-directed system (g_{k+1}). Likewise, during each memory recycle, habit strength (h) is accumulated in each second t as a function of the reward prediction errors experienced in the recycle (H is the strength accumulated up to cycle k). The probability of responding is a direct (nonlinear) function of the sum of the two strengths. Since both systems cooperate to determine responding for each second in a recycle, outcome revaluation effects depend on the proportion of strength coming from each system

single events within a recycle (response, and reinforcers obtained (or not) in each second) from which the habit strength h increases or decreases accordingly. The sum of the strength of the two systems determines the total instrumental performance for the next memory recycle. The proportion of strength of each system determines the sensitivity to devaluation.

Within the framework of Perez and Dickinson's theory (2020), a wide range of empirical data from various species is readily explained. These include not only the higher performance under ratio compared to interval training when equating reinforcement probabilities or reinforcement rates between conditions (Dickinson et al., 1983) but also the comparable rates of responding observed when subjects are overtrained and stable behavior no longer exhibits variation in reinforcement rate that could update the agent's computation of the rate correlation (Pérez et al., 2018)—the strength of the habit system is expected to be similar between conditions when reinforcement rates or probabilities are matched, resulting in similar rates of responding.

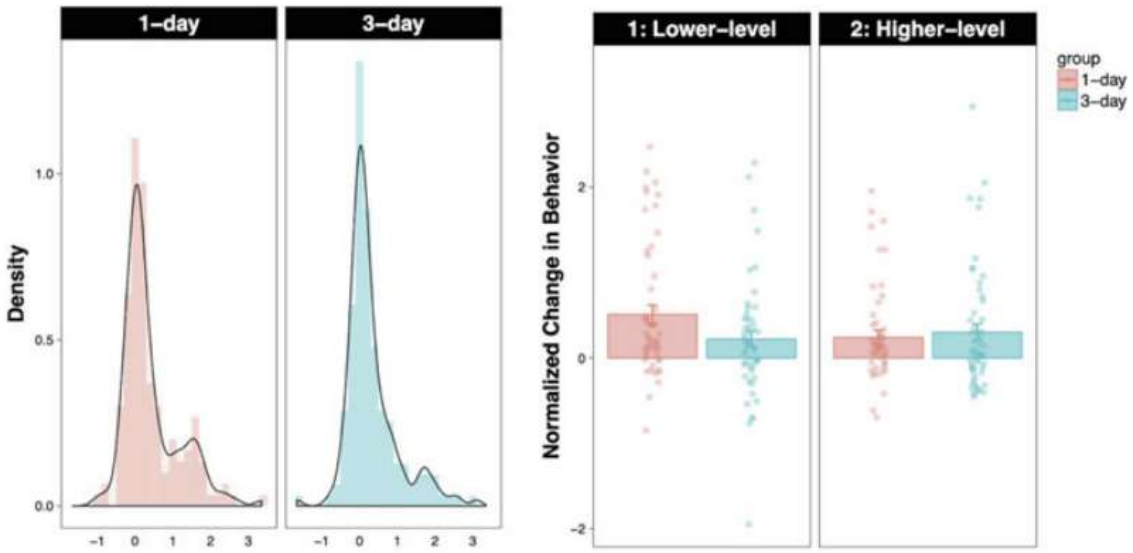

Fig. 2 Stress-affect modulates the effect of overtraining on sensitivity to devaluation. Left panel: After moderate and extended training under interval schedules, most of the participants are habitual in both groups. This is consistent with a low rate correlation established by interval schedules. The behavioral adaptation index is a measure of how goal-directed participants are. The peak at zero for most subjects shows that participants were mostly habitual in both groups. Right panel: The effect of overtraining on the transition from goal-directed behavior to habits is modulated by a combination of factors associated with stress and anxiety

This theory also predicts that behavior becomes controlled by habits and therefore insensitive to outcome revaluation after overtraining on ratio schedules, as the rate correlation weakens when responding stabilizes and variations in reinforcement rate are only weakly experienced by subjects. For interval schedules, by contrast, the transition should be more rapid.

Direct evidence supporting this prediction in humans was demonstrated recently by Pool et al. (2022), who trained participants across four different laboratories to press two keys to obtain salty or sweet rewards under two different training conditions. In the moderate training group, participants completed two sessions in a single day while those in an extensive training group participated in four sessions across three consecutive days. Following training, participants had the opportunity to consume one type of food until satiation. Liking ratings revealed that participants were indeed satiated by the procedure, as they reported a preference for the nondevalued outcome after having consumed the food.

What Pool and her colleagues found was that interval training rendered behavior habitual in most subjects in both groups (Fig. 2, left panel). Consistent with the predictions of rate-correlation theory, a weak experienced rate correlation under interval training is anticipated. Consequently, behavior should become habitual under these circumstances. However, when the authors separated participants with respect to their answers in a battery of psychological questionnaires related to stress and anxiety, they found that these factors modulated the effect of training on devaluation sensitivity (Fig. 2, right panel), in line with previous data in animals and humans demonstrating that these factors can accelerate habit formation (Schwabe

& Wolf, 2009, 2010). In Pool et al.'s data, only those participants who scored low on these factors transitioned from goal-directed to habitual control with training extension. The findings of this study indicate that individuals with high stress-anxiety scores tend to develop habitual behavior even with minimal training in the moderate group, whereas the transition occurs slower for individuals with low levels of stress/anxiety.

2 Evidence That Rate Correlation Drives Behavior in Humans

Until recently, the available data examining the specific predictions of rate correlation theory in humans, particularly concerning instrumental goal-directed performance as proposed in Perez and Dickinson's (2020) model, were limited. However, Perez and colleagues have presented compelling evidence of rate correlation's influence on human responding in two recent studies. Perez and Soto (2020) conducted an experiment in which human participants were trained to press a spacebar with the goal of flashing a triangle on the screen. Their objective was to compare the response rates generated by random-ratio (RR) and random-interval (RI) schedules for matched reward probabilities and how this performance related to causal attribution of the action-outcome contingency. Importantly, participants were not told that there were different reward schedules in effect, but simply to report the causal strength between their behavior and the triangle appearing on the screen. In spite of this lack of information, participants responded more on an RR than on an RI schedule when reward probabilities were matched within-subjects, supporting the hypothesis that instrumental performance is sensitive to rate correlation. However, if rate correlation underlies goal-directed control, then causal judgments of the action-outcome association should be higher on ratio than interval schedules given the higher rate correlation of the former. Their findings did not support this prediction. What Perez and Soto observed instead was that causal ratings followed the reward probabilities programmed by the schedules. That is, when they programmed the schedules so that participants would experience comparable reward probabilities, they observed that ratio performance surpassed interval performance (as expected), but the causal ratings were similar between schedules. In a follow-up experiment, when the experienced reward probabilities were higher than in the previous experiment, they observed again higher performance between ratio and interval conditions but also higher mean causal ratings compared to the previous experiment, suggesting that different mechanisms of causal attribution and response strength might be at play when subjects undergo free-operant training.

Perez and Soto (2020) recognized a challenge in establishing definitive conclusions regarding the impact of rate correlation on the ratio/interval difference if only RR and RI schedules were employed. The reason is that the RI schedule effectively assigns higher reward probabilities to long pauses between responses (or long inter-

Table 1 Predictions of performance for different reward schedules when reward rates or reward probabilities are matched

Reward schedule	Positive rate correlation	Differential reinforcement of long IRTs
RR	Yes	No
RI	No	Yes
RPI	No	No
Performance prediction	RR > RPI = RI	RR > RPI > RI

response times, IRTs). The more the subject waits since the last response, the more likely it is that the reinforcer will become available. It could then be argued that the lower performance of RI training is due to differential reinforcement of long IRTs, not due to the weaker rate correlation established by the schedule. For this reason, Perez and Soto implemented a random-probability interval schedule (RPI) which avoids reinforcing long IRTs by setting the reward probability for the *next* performed response according to $p = \frac{t}{Tm}$, where t is the time it has taken the subject to perform the last m IRTs and T is the scheduled interval (the inverse of the reinforcement rate). In the RPI, the reward probability is not based on the current IRT, but on the size of a number (m) of IRTs emitted before the current IRT. The RPI, in other words, considers a *local* response rate $b_m = (m + 1)/t$ given by the last $m + 1$ responses in the last t secs and varies the reward probability inversely with respect to this *local* response rate so that the agent still receives a constant reward rate of $1/T$ rewards/s independently of the size of the current IRT, establishing the null response-outcome rate correlation typical of interval schedules. Since the current IRT size contributes only a fraction $1/m$ of the change in reward probability, the RPI neutralizes the effect of timing on increasing reward probabilities for long IRTs. Therefore, an RR/RPI performance comparison should shed light on the real contribution of rate correlation to responding. Table 1 presents the predictions of rate correlation theory and IRT reinforcement on the performance under different reward schedules.

The study by Perez and Soto (2020) was not designed to test response rates for comparable reward probabilities between the schedules, but to obtain evidence of which factor was behind action-outcome causal attribution; response rates were spontaneously in line with the predictions of rate-correlation theory. To establish more definite conclusions, Perez (2021) set out to investigate instrumental performance using a within-subject design where participants experienced four different schedules and matched both reinforcement probabilities (by yoking a ratio to master interval schedules) and reinforcement rates (by yoking interval schedules to a master ratio schedule) across these options. To this end, Perez (2021) presented participants with a set of four different fictitious candy dispensers. All participants experienced a master RR schedule that was then followed by two interval schedules, an RI and an RPI (order counterbalanced across subjects), both of which were matched with respect to reinforcement rate obtained in the master RR dispenser. The purpose of comparing performance between these two interval schedules and

the master RR was to obtain direct evidence of rate correlation driving behavior (RR vs RI and RR vs RPI) and also for the role of reinforcing long pauses between responses (RPI vs RI) in the same subject. Indeed, Perez obtained evidence that performance on the RR was higher than for both of the interval schedules but that there was no effect of reinforcing pauses, in that performance was equivalent between the two interval schedules. Perez then matched the mean reward probability experienced by subjects in the two interval schedules to a final RR schedule, to test if controlling for reward probability but not reinforcement rate would change the results. Again, RR performance was higher than both RI and RPI schedules. Moreover, there was no difference in performance between these two interval schedules, indicating a null contribution of IRT reinforcement on responding. Given the limited training participants underwent for multiple options, Perez's results suggest that rate correlation drives goal-directed performance in humans.

3 Rate-Correlation Theory and Avoidance

Perez and Dickinson (2020)'s model was concerned with the most common experimental procedure: appetitive conditioning. However, in their paper they also speculated that the same ideas may be extended to avoidance behavior where responding is contingent to a reduction in the rate of an aversive event. Although the idea that avoidance behavior can be goal-directed is not new (Seligman & Johnston, 1973), it was not until a set of experiments some 10 years ago came out that the idea regained traction (Fernando et al., 2014a, b). Apart from a single study by Wang et al. (2018) using discrete-trial procedures with a human RL framework, no experiments had suggested a role for multiple systems in explaining avoidance, and indeed, RL theory had been so far based on the assumption that a single model-free system was sufficient to explain it (see Maia, 2010).

However, in a series of experiments, Fernando et al. (2014a, b) have reported compelling evidence demonstrating that goal-directed and habitual processes also underlie avoidance behavior. They employed a free-operant schedule and revaluation procedure to investigate the impact of aversive outcome devaluation. Rats were trained to engage in lever-pressing to avoid foot-shocks delivered at fixed intervals under a variable-cycle (VC) schedule. During an extinction test, a group of rats that received noncontingent shock presentations while under the influence of morphine exhibited decreased responding compared to a control group that did not receive morphine. These findings demonstrate that diminishing pain and devaluing the aversive nature of the shock yield similar effects to the devaluation procedures employed in appetitive conditioning, revealing that goal-directed behavior can be performed to avoid undesirable outcomes, just as it can be performed to obtain pleasurable outcomes.

That avoidance is controlled by more than one behavioral system was demonstrated in another experimental study by Fernando et al. (2014b), who set out to

examine the role of habit learning in free-operant avoidance. To explain avoidance, the literature usually distinguishes between events prior and after the response as drivers of avoidance responding. In signaled paradigms, where a stimulus is presented before an impending shock, this warning signal produces fear as it predicts the unpleasant outcome; the avoidance response releases the fear produced by the signal with a subsequent period without shocks, called the safety period. But there is also evidence that feedback after or during an avoidance response can produce safety.

In Pavlovian conditioning, stimuli can signal the presence or absence of an event, which could be appetitive or aversive (Konorski, 1967). When a stimulus signals the absence of an otherwise present outcome by being inversely correlated with it, the stimulus turns into a conditioned inhibitor. When the absence is that of an appetitive event, the inhibitor acquires specific properties, such as retarding learning about other appetitive outcomes or interacting with the learning accrued by other stimuli (Dickinson & Pearce, 1977; Rescorla & Wagner, 1972; Wagner & Rescorla, 1972; Williams & McDevitt, 2002). By contrast, when the absence is that of an aversive event, the stimulus acquires the opposite properties. For example, including an explicit stimulus after an avoidance response can turn the stimulus into an aversive conditioned inhibitor, because it arranges for a negative Pavlovian relationship between the stimulus and the unpleasant outcome. When this happens, the stimulus can produce motivation and become a conditioned reinforcer, in that animals would work for its presentation in the absence of the primary reinforcer.

Consistent with this notion, Fernando et al. (2014b) observed that the addition of a feedback stimulus to the response resulted in augmented avoidance responding in a free-operant schedule. To assess the type of behavioral control exerted by the stimulus, they also included an outcome revaluation test where presentations of the stimulus were associated with morphine. Their findings indicated that this enhancement of responding did not persist in an outcome revaluation test, suggesting that responding for the conditioned reinforcer was under habitual control. Collectively, these two studies by Fernando and colleagues provide the first evidence for the involvement of both goal-directed and habitual systems in avoidance learning.

Based on these findings, Perez and Dickinson (2023) have recently proposed a dual-system model of avoidance that includes a rate-correlation system that encodes specifically aversive events in free-operant avoidance schedules. In this type of schedule, shocks are programmed to occur at random intervals (a variable cycle, VC, in avoidance terminology), while subjects have the freedom to respond at any time. If a subject responds before the next scheduled shock, the shock is canceled, and a new cycle begins. Importantly, all responses made between the current time and the avoided shock have no impact on subsequent programmed shocks; they are irrelevant to the environment, except for the reinforcing proprioceptive feedback stimuli that shape habitual responses.

The only study to systematically investigate avoidance using different VC schedules was conducted by de Villiers (1974). Using parameters similar to those employed by Fernando et al. (2014a), de Villiers showed that the response rate of

his rats consistently decreased as the interval between shocks (the shock-shock) increased. To investigate if a rate-correlation system could be applied to this free-operant avoidance setting, Perez and Dickinson (2023) modeled a short-term memory where the agent computes the rate correlation between the responses per memory sample and received shocks per memory sample. The memory was recycled so that one sample was randomly forgotten by the agent and responding in a new sample was determined by a sublinear combination of the current experienced correlation between responses and shocks per sample and the average correlation experienced during training. Response strength from the goal-directed system, g, is explained in their model under this framework by noting that the correlation between responding and shocks in memory r is negative, but so is the incentive value of the outcome I. Therefore, as $g = Ir$ (the incentive value I times the experienced rate correlation r), the system yields a positive response strength. By simulation, Perez and Dickinson (2023) demonstrated that a rate-correlation system can capture the qualitative relationship between response rate and shock rate reduction observed by de Villiers (1974). These results suggest that rate correlation theory offers a plausible explanation for performance not only in the context of appetitive rewards but also in aversive outcomes such as shocks. Considering that the parameters simulated by Perez and Dickinson (2023) align closely with those utilized by Fernando et al. in their recent experiments, it is likely that goal-directed control also underlies the behavior observed in de Villiers' (1974) rats.

More problematic for a dual-system view of avoidance in free-operant training is formulating the cooperation between the goal-directed strength g and the habit strength h. Perez and Dickinson (2023) pointed out the difficulties of determining h from a model based on reinforcement by conditioned inhibition for feedback stimuli, as there is no sufficient evidence to provide a full Pavlovian theory of inhibitory signals in avoidance. Perez and Dickinson (2023) speculated, however, that the prediction error of the habit system could be determined by the difference in the strength of Pavlovian inhibition elicited by the feedback stimuli at the time when a response is performed and the current habit strength. Likewise, from a psychological perspective, it is not clear how the predictions for the habit strength should incorporate the predictions of the goal-directed system, as in their original theory. The reason for this is that, unlike their original model for appetitive conditioning where both systems were driven by a common outcome, in the case of avoidance the events that drive each system are different. For the goal-directed system, it is the shock reduction, but for the habit system, it is the feedback stimuli. These are pending issues that will require more data to be resolved.

Regardless of these pending issues, their approach shows how free-operant avoidance can be captured by a rate-correlation goal-directed system and a habitual system based on the inhibitory properties of the response-generated feedback stimuli, both of which should cooperate to explain responding and devaluation sensitivity.

4 Extinction, Causal Degradation, and Goal-Directed Control

In a few seminal experiments in rats, Prof. Rescorla demonstrated that extinction procedures—where outcome delivery is suspended—did not affect sensitivity to outcome devaluation. In an elegant design, Rescorla (1993) trained rats to perform two different actions for two different outcomes before an extinction session where the outcome was suspended for one of the actions. Then, to recover responding of the extinguished actions, he introduced another session where a different outcome was delivered contingent upon responding. Finally, he devalued one of the outcomes and compared responding on the extinguished and nonextinguished actions by having equated the response rates in the previous retraining session. Rescorla found that extinguished actions were equally sensitive to devaluation as non-extinguished actions, showing that goal-directed control survives extinction.

The fact that an action that no longer produces an outcome can still be sensitive to changes in the value of that outcome is a puzzling and challenging phenomenon from a theoretical perspective. If the action-outcome connection has been severed by extinction, then it is unclear why there should still be any connection between the action and the incentive value or utility of the outcome. So far there is only one explanation for this result. Perez and Dickinson (2020) propose that the habit system inhibits the contribution of the goal-directed system to responding during extinction. As noted before, their theory assumes that both a correlational-based system for goal-directed control and a habitual system based on reward prediction error cooperate to produce response strength. This prediction error is defined as the difference between the current habit strength and the sum of the response strength of the goal-directed and habitual systems. For limited amounts of training, the habitual system is not strong enough to control responding; the correlational-based goal-directed system is positive and explains most of the behavior. When extinction comes on, the outcome is suspended and the short-term memory system does not contain any events to compute a rate correlation, so g remains at the same level as the average rate correlation experienced during the task. However, the habit system, whose prediction includes the goal-directed strength g, continues experiencing a negative reward prediction error driven by the positive and constant value of g across extinction. This implies that eventually, in future memory recycles, h will become negative, counteracting the effect of g on responding and explaining why behavior extinguishes with sufficient extinction training. However, since outcome devaluation is determined by g, goal-directed responding is still active after extinction. As can be appreciated, this result is explained in Perez and Dickinson's model by appealing to an inhibitory process, whereby the habit system masks the contribution of the rate-correlation system to responding.

An additional prediction is anticipated from Perez and Dickinson's theory regarding a classic manipulation of the action-outcome link. When the causal action-outcome contingency is compromised by delivering outcomes at the same rate as in training but independently of responding, the experienced rate correlation

within memory weakens, leading the goal-directed system to systematically reduce its contribution to overall responding (Balleine & Dickinson, 1998; Vaghi et al., 2019). In contrast, the habit strength h takes longer to be influenced by the negative prediction errors as well as by reinforcement occurring by chance due to some responses being reinforced. Consequently, devaluing the outcome following such manipulation renders behavior insensitive to devaluation, as the habit strength h surpasses the goal-directed strength g when the outcome is devalued. Crimmins et al. (2022) provide evidence for this prediction in rats pressing levers.

In two experiments utilizing a 2-lever 2-outcome design, Crimmins et al. first observed that causal degradation did not affect goal-directed behavior. Rats continued to decrease their response rate for a devalued outcome compared to a valued outcome even after causal degradation, just as in the case of Rescorla's experiments on extinction. However, their design did not rule out the possibility that the diminution in responding was a consequence of the Pavlovian association between the stimuli associated with pressing the different levers and the rewarding outcome, which has previously been demonstrated to affect responding selectively and independently of the instrumental relationship between actions and outcomes. When the influence of Pavlovian stimuli in motivating the performance of the still-valued action was neutralized by employing a bidirectional vertical pole—which effectively neutralized the subjects' experience with the surrounding stimuli regardless of the action performed—the devaluation effect vanished and rats did not decrease their responding to the devalued compared to the valued action. Consistent with the predictions of Perez and Dickinson's theory, rats became insensitive to outcome revaluation.

In a nonpublished study, Perez et al. (in preparation) have extended this work by investigating the impact of different manipulations of action-outcome contingency on human goal-directed actions. In their design, online participants were required to perform four specific actions (keypresses) in order to obtain two different fictitious outcomes: silver and gold coins (Gillan et al., 2015). Each pair of actions was concurrently trained within individual blocks of training. The outcomes were programmed to be delivered under an RI-7s schedule and a changeover delay was imposed so that switching between options was not reinforced. In one of their experiments, after training the pairs of actions concurrently, they selectively extinguished one pair by suspending the delivery of the associated outcomes, while the other pair did not undergo extinction. Subsequently, they retrained the extinguished actions with a different type of coin to restore the response levels observed at the end of the initial training. To devalue the outcome, participants were informed that the fictitious piggy banks where one of the earned outcomes was being deposited throughout the experiment had become full. This manipulation effectively reduced the probability of participants selecting that particular coin during a subsequent free consumption test, where they had the opportunity to collect coins of all types from the screen, providing evidence that the devaluation was successful in decreasing the value of the outcome relative to the other, nondevalued outcome. Lastly, participants underwent a final test in which the pairs of actions presented during training were available, but the outcome was hidden behind a

curtain, preventing participants from seeing the earned rewards. This effectively constituted a "pseudoextinction" test, because participants still thought the rewards were coming up behind the curtain as during training. During the test phase, participants exhibited a consistent reduction in responding to the devalued outcome across both extinguished and nonextinguished actions. This decrease in responding was observed in both conditions, providing a replication of Rescorla's findings and aligning with the predictions of the rate-correlation system proposed by Perez and Dickinson (2020).

The authors also ran another experiment where the action-outcome contingency was partially degraded by including on top of the RI schedules free delivery of outcomes of each type in each individual block of training of two concurrent actions. Although the manipulation should weaken the causality between action and outcome compared to regular RI training, participants were still goal-directed during the last test, suggesting that goal-directed control could also survive partial degradation of the causal action-outcome contingency. It remains to be tested if full degradation—where outcomes are delivered freely independently of responding—would also keep goal-directed behavior intact. If rate correlation is driving goal-directed behavior, full degradation should render behavior insensitive to devaluation, just as in the Crimmins et al. (2022) experiment.

Overall, these results support the predictions of Perez and Dickinson's model, highlighting the role of rate correlation and reinforcing the understanding of the different dynamics between goal-directed and habitual systems in response to causal degradation and outcome devaluation.

5 Future Extensions and Implications for Other Areas of Research

The key message conveyed in this chapter is that a substantial body of recent evidence, encompassing both human and animal studies, aligns with the predictions of a dual-system framework in which goal-directed behavior is driven by the mechanism of rate-correlation theory. This theory provides an explanation for the differences in instrumental performance between different reinforcement schedules in humans and the susceptibility to outcome devaluation following schedule training, extinction, and causal degradation in both rats (Crimmins et al., 2022) and humans (Perez et al., in preparation). Previously overlooked areas of research, such as exploring avoidance behavior from this perspective, have regained attention and opened new avenues of investigation.

One still-pending issue with respect to rate correlation and goal-directed behavior is the challenge to produce habitual behavior in the laboratory. After the first demonstration of habitual behavior in humans by manipulating training extension by Tricomi et al. (2009) inside an MRI scanner, follow-up attempts have been largely unsuccessful (de Wit et al., 2018). The demonstration of Pool et al. (2022)

of habitual control for both moderate and extensive training is consistent with rate-correlation theory, as the interval schedules employed in their study are predicted to have weak goal-directed strength given the low rate correlation experienced even for moderate amounts of training. As noted, most of their participants only transitioned to habits with training when their self-reported levels of stress/anxiety were low, whereas those reporting high levels showed habits from the outset of the experiment. This shows that habitual control can be observed in humans in the laboratory, but it does not explain the discrepancy between Tricomi et al.'s (2009) study and the failures to replicate it.

A logical progression from this experiment would involve incorporating a ratio schedule instead of an interval schedule during the training phase. As suggested by Perez and Dickinson (2020), it is ratio schedule that demonstrates initial positive rate correlations, gradually diminishing as behavior stabilizes. While this presents a testable and unambiguous hypothesis, the practical implementation faces challenges due to variations in individual participants and slight disparities in manipulanda, instructions, and experimental procedures across different laboratories. For instance, the ease of executing keypresses and the absence of an explicit cost for exceeding the required responses under interval training might have hindered participants' comprehension of the interval contingency and their sensitivity to it. To address this concern, a potential solution would involve employing a manipulandum with some resistance, which not only introduces an explicit cost but also imposes an evident energetic burden on participants who engage in excessive responding. This adjustment would likely enhance their engagement with the reinforcement schedule.

The investigation of the interplay between behavioral systems in free-operant training and discrete-trial scenarios within human reinforcement learning (RL) models (model-free and model-based) represents another important area of inquiry that has not yet been explored. Model-free computations have been associated with habitual strategies, while model-based computations, which consider future reward probabilities, are believed to underlie goal-directed control. Prior evidence indicates that both systems can exert influence on behavior and are correlated with distinct computational algorithms in the brain. However, no studies have yet examined whether goal-directed behavior, as evidenced by the experienced rate correlation, correlates with the degree of reliance on model-based strategies (O'Doherty et al., 2021). This would require using free-operant and discrete-based choice tasks in the same subjects. Such design should help elucidate the connection between the psychological construct of habit and the algorithmic-based explanation from reinforcement learning theory.

Likewise, it remains unclear whether the same brain structures would exhibit relatively greater activation for both behavioral strategies at the individual level. By conducting such studies, valuable insights could be gained regarding whether distinct computational strategies underlie a shared behavioral system. This would suggest that the brain is selecting between these strategies at a meta-level, deploying them based on the specific scenario being encountered. These results would shed light on the intricate interplay between computational processes and behavior,

unraveling the mechanisms by which the brain navigates and adapts to different situations.

Rate-correlation theory is also readily applicable to other natural scenarios. In the context of foraging theory (which focuses on how animals search and utilize energy resources), the classification of resources as depleting or nondepleting parallels the concepts of interval and ratio schedules, respectively. Foraging theory (Hayden, 2018; Stephens & Krebs, 1986) addresses the problem of energy intake from these resources by applying microeconomic principles, where organisms maximize expected utility by considering the costs and benefits of resource acquisition, including risk and competition. This cost-benefit analysis enables organisms to make optimal decisions in resource utilization. Generally, nondepleting resources, characterized by predictability and sustainability, promote more consistent foraging behavior (higher response rate), while the dynamic nature of depleting resources leads to greater variability in foraging strategies (lower response rate).

Despite foraging theory's foundation in expected utility maximization, it lacks a correlational mechanism to explain the higher behavioral consistency observed with nondepleting resources compared to depleting ones. However, it is possible that even when resources share similar probabilities or rates, nondepleting resources still elicit a greater degree of foraging, indicating that animals in natural environments might be considering longer time periods in their decision-making beyond the emphasis on reward probability per each patch visit. Incorporating dual-system theory, specifically rate-correlation theory, alongside foraging theory's utility-maximizing agent, may offer a more comprehensive explanation by integrating both goal-directed and habitual processes in foraging behavior.

Indirect evidence suggests that rate correlation may also play a role in human behavior beyond laboratory settings. In consumer behavior, for instance, it has been observed that incentives designed to stimulate product demand are inversely related to the rate of consumer engagement. Specifically, individuals with lower purchase rates baseline levels (response rate) for a product are more likely to increase their purchases when promotions, incentives, or loyalty programs are introduced, as they are still experiencing a positive correlation between the rate of purchases and the rewards. However, as purchase rates increase, the behavioral variability decreases, leading to a shift towards habitual consumption. As these programs are specifically aimed at emphasizing the attributes of the products (their sensory properties), they should not be effective in boosting demand when purchase rate is high (Taylor & Neslin, 2005). Additionally, consumers who selectively choose products often exhibit reduced loyalty to specific stores, aligning with the notion that constantly switching between products impedes habitual consumption by a constant reexperience of different rewards, which maintains a positive experienced correlation between the response and the rewards (Fox & Semple, 2002). This corresponds to the idea that multiple actions and outcomes contribute to goal-directed behavior, which maintains consumers' attention to the different attributes of products (Kosaki & Dickinson, 2010). For these consumers, loyalty programs or promotions should be expected to have a significant impact on their purchasing behavior.

Whatever the extensions of rate-correlation theory, the evidence in the last decade or so supports Baum's (1973) original conception of rate correlation as a promising framework for explaining a diverse range of goal-directed decisions observed in both laboratory settings and, potentially, our everyday lives. Indeed, the unsignaled nature of behavior in free-operant settings is closer to the situations subjects encounter in real life than the discrete-trial procedures generally used in human laboratory experiments. The scenarios modeled by free-operant training may exhibit superior ecological validity compared to those typically explored in human reinforcement learning literature, as they mimic the spontaneous performance of an action in natural environments which is not signaled by specific discrete stimuli and more readily attributed by subjects to a general situation or stimulus configuration. Perez and Dickinson's (2020) theory offers an explanation for many of the free-operant phenomena found in the laboratory concerning actions and habits, but extending the idea to real-life data would represent a significant advancement in this area. Some researchers have undertaken such an approach, but providing formal explanations from a computational model of actions and habits in real life would provide us with valuable insights about the mechanisms governing actions and habits in society.

References

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology, 33*(2), 109–121. https://doi.org/10.1080/14640748108400816

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37*(4–5), 407–419.

Baum, W. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior, 20*, 137–153.

Baum, W. (2012). Rethinking reinforcement: Allocation, induction, and contingency. *Journal of the Experimental Analysis of Behavior, 97*(1), 101–124. https://doi.org/10.1901/jeab.2012.97-101

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58*(5), 313–323. https://doi.org/10.1037/h0054388

Crimmins, B., Burton, T. J., McNulty, M., Laurent, V., Hart, G., & Balleine, B. W. (2022). Response-independent outcome presentations weaken the instrumental response-outcome association. *Journal of Experimental Psychology: Animal Learning and Cognition, 48*(4), 396.

de Villiers, P. A. (1974). The Law of Effect and avoidance: A Quantitative Relationship Between Response Rate and Shock-Frequency Reduction. *Journal of the Experimental Analysis of Behavior, 21*(2), 223–235. https://doi.org/10.1901/jeab.1974.21-223

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General, 147*(7), 1043. https://doi.org/10.1037/xge0000402

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 308*(1135), 67–78.

Dickinson, A., & Pearce, J. M. (1977). Inhibitory interactions between appetitive and aversive stimuli. *Psychological Bulletin, 84*(4), 690–711. https://doi.org/10.1037/0033-2909.84.4.690

Dickinson, A., Nicholas, D. J. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology, 35*(789759670), 35–51. https://doi.org/10.1080/14640748308400912

Fernando, A., Urcelay, G., Mar, A., Dickinson, A., & Robbins, T. (2014a). Free-operant avoidance behavior by rats after reinforcer revaluation using opioid agonists and D-amphetamine. *Journal of Neuroscience, 34*(18), 6286–6293.

Fernando, A., Urcelay, G. P., Mar, A. C., Dickinson, A., & Robbins, T. W. (2014b). Safety signals as instrumental reinforcers during free-operant avoidance. *Learning & Memory, 21*(9), 488–497.

Fox, E. J., & Semple, J. (2002). *Understanding "cherry pickers:" How retail customers split their shopping baskets* (Unpublished manuscript). Cox School of Business, Southern Methodist University.

Gillan, C. M. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience, 15*(3), 523–536. https://doi.org/10.3758/s13415-015-0347-6

Gremel, C. M., & Costa, R. M. (2013). Orbifrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications, 4*(May), 1–12. https://doi.org/10.1038/ncomms3264

Hayden, B. Y. (2018). Economic choice: The foraging perspective. *Current Opinion in Behavioral Sciences, 24*, 1–6. https://doi.org/10.1016/j.cobeha.2017.12.002

Konorski, J. (1967). Integrative activity of the brain. University of Chicago Press: Chicago.

Kosaki, Y., & Dickinson, A. (2010). Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 36*(3), 334–342. https://doi.org/10.1037/a0016887

Maia, T. V. (2010). Two-factor theory, the actor-critic model, and conditioned avoidance. *Learning & Behavior, 38*(1), 50–67. https://doi.org/10.3758/LB.38.1.50

O'Doherty, J. P., Lee, S. W., Tadayonnejad, R., Cockburn, J., Iigaya, K., & Charpentier, C. J. (2021). Why and how the brain weights contributions from a mixture of experts. *Neuroscience & Biobehavioral Reviews, 123*, 14–23.

Perez, Oh, Rojas, & Merlo. (in preparation). *The causal status of goal-directed actions after manipulations of the action-outcome association*.

Perez, O. D. (2021). Instrumental behavior in humans is sensitive to the correlation between response rate and reward rate. *Psychonomic Bulletin & Review, 28*(2), 649–656.

Perez, O. D., & Dickinson, A. (2020). A theory of actions and habits: The interaction of rate correlation and contiguity systems in free-operant behavior. *Psychological Review, 127*(6), 945–971. https://doi.org/10.1037/rev0000201

Perez, O. D., & Dickinson, A. (2023). Dual-system avoidance: Extension of a theory. *BioRxiv*. https://doi.org/10.1101/2023.05.24.542134

Perez, O. D., & Soto, F. (2020). Evidence for a dissociation between causal beliefs and instrumental actions. *The Quarterly Journal of Experimental Psychology, 73*(4), 495–503.

Pérez, O. D., Aitken, M. R. F. F., Milton, A. L., Dickinson, A., Perez, O. D., Milton, A. L., Aitken, M. R. F. F., & Dickinson, A. (2018). A re-examination of responding on ratio and regulated-probability interval schedules. *Learning and Motivation, 64*, 1–8. https://doi.org/10.1016/j.lmot.2018.07.003

Pool, E. R., Gera, R., Fransen, A., Perez, O. D., Cremer, A., Aleksic, M., Tanwisuth, S., Quail, S., Ceceli, A. O., & Manfredi, D. A. (2022). Determining the effects of training duration on the behavioral expression of habitual control in humans: A multilaboratory investigation. *Learning & Memory, 29*(1), 16–28.

Rescorla, R. A. (1993). Preservation of response-outcome associations through extinction. *Animal Learning & Behavior, 21*(3), 238–245. https://doi.org/10.3758/BF03197988

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In W. F. Black & A. H. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts. https://doi.org10.1037/a0030892

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience, 29*(22), 7191–7198.

Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology, 35*(7), 977–986.

Seligman, M. E., & Johnston, J. C. (1973). A cognitive theory of avoidance learning. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning* (pp. 69–110). Winston & Sons Inc.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory* (Vol. 6). Princeton University Press.

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

Taylor, G. A., & Neslin, S. A. (2005). The current and future sales impact of a retail frequency reward program. *Journal of Retailing, 81*(4), 293–305.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *The European Journal of Neuroscience, 29*(11), 2225–2232. https://doi.org/10.1111/j.1460-9568.2009.06796.x

Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2019). Action-outcome knowledge dissociates from behavior in obsessive-compulsive disorder following contingency degradation. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4*, 200–209.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and learning*. New York: Academic.

Wang, O., Lee, S. W., O'Doherty, J., Seymour, B., & Yoshida, W. (2018). Model-based and model-free pain avoidance learning. *Brain and Neuroscience Advances, 2*, 239821281877296. https://doi.org/10.1177/2398212818772964

Williams, B. A., & McDevitt, M. A. (2002). Inhibition and superconditioning. *Psychological Science, 13*(5), 454–459.