Common Principles Underlying Models of Sequential Effects

Matt Wilder

Abstract

Human behavior fluctuates—an appropriate action at one moment in time may not be appropriate at the next moment. The study of sequential effects seeks to explicate the dependence these fluctuations have on the exact sequence of recent events. For example, in a simple choice experiment, how will an individual's speed of response vary as a function of the sequence of previous stimuli? Though often studied in controlled experiments, sequential effects are not just a curiosity of the laboratory and in fact arise in many real world settings. They have been identified in automobile driving, legal reasoning, jury evidence interpretation, clinical assessments, financial decisions, and basketball players' shot choice. Certainly it seems that if you look for sequential effects in any domain, you find them. Because of this widespread presence, there are numerous theoretical and computational accounts. Though some synthesis has been offered within particular domains, the field is lacking an organization that spans the entire literature. The goal of this work is to bring clarity to this scattered field by presenting a few computational-level principles that characterize most models of sequential effects. By focusing on a computational level of analysis, we aim to uncover the primary purpose of sequential effects in cognition (i.e., *why* they exist instead of *how* they are produced). We present three main computational explanations for sequential effects: (1) they reflect the attempt to predict future events in a dynamic environment; (2) they result from the need to continually adjust the sensitivity of perceptual and decisional processing; and (3) they reflect the generalization of knowledge from the recent past to new experiences. In addition, we identify two general classes that encompass many mechanistic models which lack a clear computational perspective: neural inertia models and incremental learning models. In the process of categorizing these models, we highlight a strong connection between sequential effects and adaptation across all explanations. Furthermore, we suggest the potential value to be gained in the latter two computational explanations by rooting models in normative principles of adaptation as has been done with some models classified in the first computational explanation. We believe that our theoretical perspective will provide important guidance in interpreting and understanding sequential effects, developing new models, and exploring the role of adaptation across all domains of cognition.

Introduction

Our world is structured such that we experience a steady stream of sensations and continually formulate decisions and actions in response to them. Understanding this input/output relationship is the primary goal of psychology. Generally research is focused on characterizing the response behavior to a particular sensation or a small set of sensations. Consider the typical experimental design where subjects are exposed to some sort of stimulus on a given trial and are required to make an appropriate response. The goal usually is to quantify the difference in mean behavior across several stimuli. However, there is a significant flaw in this methodology. Humans do not produce independent responses to each subsequent experience; instead our behavior is deeply influenced by the context of the moment. The study of sequential effects in behavior recognizes this reality and aims to develop a more complete understanding of how individuals behave when encountering different stimuli and how this behavior is modulated by the current context (i.e., the sequence of sensations, decisions, and actions that precede the response under scrutiny).

Sequential effects have long received attention in the experimental literature, though they are acknowledged and addressed in only a small minority of studies. Despite the fact that sequential effects are ignored in most analyses, experimental evidence suggests that they are present in nearly all aspects of cognition and are even ubiquitous across more rudimentary components of our nervous system such as motor control. In most cases, when investigators look for sequential effects, they find them.

One early example of a sequential effect is the gambler's fallacy identified by Jarvik (1946) in which after a long sequence of repeated events, individuals have a bias towards expecting a reversal of the sequence even if the events are independent and equally probable. This bias resulted in a huge win by a casino in Monte Carlo one evening in August 1913 when one of the roulette wheels logged a remarkable 15 blacks in a row, sparking a rush of bets on red, and then continued to a record streak of 26 consecutive blacks, breaking the bank of many individuals who kept doubling down their bets believing that there was no possible chance the streak could continue.

Sequential effects have been demonstrated in a wide range of experimental paradigms including simple choice tasks (e.g., Bertelson, 1961; Hyman, 1953; Remington, 1969), probability matching (e.g., Estes, 1957), absolute judgments (e.g., Garner, 1953; Ward & Lockhead, 1970), magnitude estimation (e.g., Jesteadt et al., 1977; Ward, 1973), categorization (e.g., Petzold, 1981; Treisman & Williams, 1984), visual search (e.g., Chun & Jiang, 1998; Maljkovik & Nakayama, 1994), and language production (Bock & Griffin, 2000).

Beyond these somewhat contrived laboratory tasks, sequential effects have been identified in significant real world situations. For example, recent braking or acceleration actions of automobile drivers can explain variability in response latencies of up to 100 ms, potentially the difference between a collision and a near miss (Doshi et al., 2012). Professional basketball players' choice of shot location has been shown to depend directly on recent attempts and successes (Neiman & Loewenstein, 2011). A bias of the recent past on decision-making has been demonstrated in legal reasoning and jury evidence interpretation (e.g., Furnham, 1986; Hogarth & Einhorn, 1992), clinical assessments (Mumma & Wilson, 2006) and financial decisions (Johnson & Tellis, 2005; Vlaev et al. 2007).

Furthermore, sequential effects are present in motor control (Sheidt et al., 2001) and pain sensation (Mozer et al., 2011), suggesting that dependencies on the recent past are not just a property of high-level cognition. In fact, other animals have been shown to exhibit sequential dependencies. For example, rats adjust their behavior according to the sequence of rewards (Gallistel et al., 2001), the foraging behavior of starlings is biased by recent experience (Cuthill et al., 1990), and the flower choice of bumblebees seeking nectar is influenced by the few recent flower visits (Real, 1991). Figure 1 presents several examples of sequential effects in different domains.

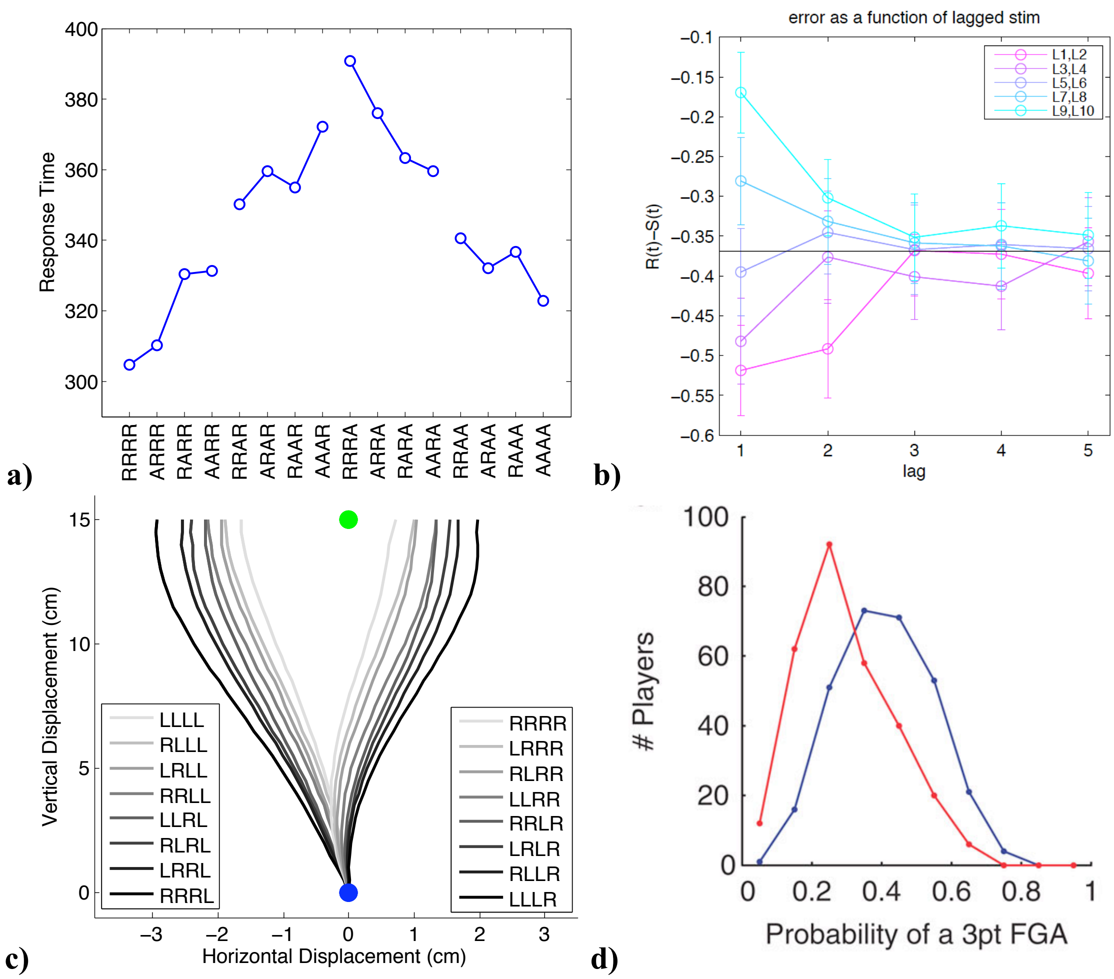

Figure 1. Examples of several different kinds of sequential effects. a) Variation in response time (RT) in a two-alternative forced-choice task as a function of the 4-back history from Cho et al. (2002). RTs are slower when the current trial (top of sequence label) does not match the recent trials. b) Judgment error in a line estimation task as a function the stimulus presented a different lags in the past. Short lines (L1 and L2) that occur in the recent past have a negative bias on the mean response for the current trial. Similarly, long lines (L9 and L10) have a positive bias on mean response. This effect is an assimilation effect because judgments are pushed in the direction of recent stimuli. Note also that the effect is stronger at more recent lags. c) Mean trajectories in a motor control study with perpendicular perturbations. Errors are smaller when the current perturbation is in the same direction (right or left) as previous perturbations. d) Sequential effects in basketball players' shot choice (Neiman & Loewenstein, 2011). Histogram of players' probability of selecting a 3pt shot after made (blue) and missed (red) shots. Choosing a 3pt shot is more probable after

a made shot. Though not depicted in this plot, the authors also show that shot choice is biased by more than just the previous trial.

The diversity in these findings has led to a proliferation of theoretical models that seek to explain sequential effects. However, in many cases, researchers view sequential effects as a quirk of their specific domain and account for them by making subtle tweaks their domain specific models in a way that obscures the common principles underlying sequential effects. Our goal in this work is to survey the wide swath of theoretical and computational models and whittle this disparate collection down to a few central explanations for why sequential effects occur. By extracting the core principles from this scattered literature, we hope to clarify the role of sequential effects in behavior and provide guidance for interpreting sequential effects across all domains.

In synthesizing these models, we find that in most cases, sequential effects can be understood as an efficient solution to the challenges imposed by a dynamic environment under the constraints of a limited cognitive architecture. This perspective highlights the importance of understanding and recognizing sequential effects in cognition. Instead of taking the common view that these effects are a suboptimal idiosyncrasy of the brain, we encourage researchers to explore how sequential effects can further enrich our understanding of all aspects of behavior. For example, our analysis finds a deep connection between sequential effects, adaptation, learning, memory, and the statistics of the environment. In fact, sequential effects may be a manifestation of the brain's attempt to keep up with continually changing environmental conditions and may serve as the initial stage in developing more complex adaptive behavior.

Neural Inertia

Most early explanations for sequential effects have a strong mechanistic feel, attributing them to residual neural activity that yields a simple form of priming. The general idea is that the previous trial or sequence of trials produces a residual trace of neural activity or an implicit short-term memory that influences subsequent behavior. Figure 2 depicts a caricature of a neural inertia model. In the early literature, residual neural activity is often associated with a *facilitation* effect where performance is enhanced on repetition trials—where the current stimulus is the same as the previous stimulus (Bertelson, 1961; Hyman, 1953). This priming is not just limited to the relationship between the current and previous trial but can extend several trials into the past with greater facilitation after a string of repetitions (Remington, 1969). This is closely linked to the notion of *expectancy* in early theoretical accounts, where subjects are assumed to adjust behavior according to a loose implicit expectation about what will occur next formed from recent experience. Falmagne (1965) presents an early formalization of this concept in a model for a multi-choice reaction time task. Though presented via a slightly more complex mathematical framework, the model essentially implements a decaying memory trace for each potential stimulus and predicts response time (RT) for a given stimulus to be a function of an individual's level of *preparation*, defined as the strength of the trace for that stimulus. Typically, expectancy refers to a positive bias towards recent experience. However, the effect can also be negative as in Jarvik (1946), though the gambler's fallacy most likely results from explicit memory rather than implicit short-term memory.

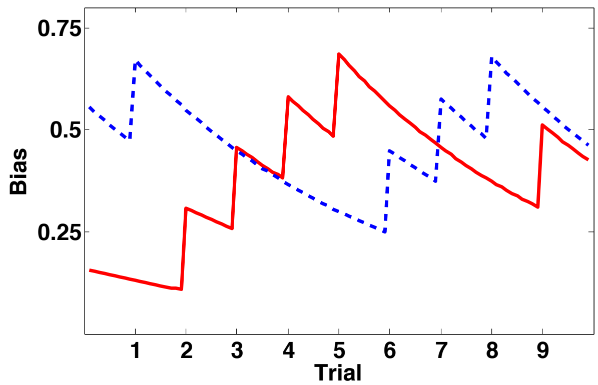

Figure 2. An example of neural inertia. If the level of bias affects behavior in some way, there will be spillover of neural activity from previous trials that will yield sequential effects in behavior.

The general theoretical perspective underlying these neural inertia accounts is that the cognitive system has a built in expectation for successive events to be similar. In this way, sequential effects can be viewed as a simple heuristic to improve performance in situations that do in fact exhibit positive autocorrelation from one trial to the next.

Much of the subsequent theoretical work on choice RT tasks targets subtleties in the experimental data such as how behavior changes with the response-stimulus interval (RSI), the stimulus/response mappings, different baserate probabilities, sequential autocorrelation structure, the number of stimuli or choices, and effects of practice (Hale, 1969; Kirby, 1976; Laming, 1968; Soetens et al., 1984; Soetens et al, 1985; Vervaeck & Boer, 1980). The theories put forth appear diverse on the surface, with conflicting perspectives on the presence and role of facilitation and expectancy. However, in most cases the sequential effects still result from some form of residual neural activity occurring within the mechanisms proposed. Much of the debate focuses not on whether neural inertia is present, but rather what is the time course of decay and how the different mechanisms combine.

For example, Vervaeck and Boer (1980) explain sequential effects by a variable state of excitation or inhibition of two processing pathways, one for each choice. When a pathway is used on a trial, it experiences an increase in excitation that leads to faster responses when that pathway is used on subsequent trials. Further, they propose that the state of the unused pathway is differentially affected by repetitions and alternations in a way that depends on RSI and practice. Soetens et al. (1985) conclude that facilitation and expectancy are completely different mechanisms that predict different patterns of sequential effects. The extent to which each mechanism affects performance depends importantly on experimental design, especially RSI and response compatibility. The authors explain facilitation via rapidly decaying memory traces but are less clear as to how expectancy is developed. It is suggested that expectancy loosely corresponds to the local impression from a run of binary stimuli which could naturally be derived from a decaying short-term memory of recent trials.

A more recent class of models that seek to explain sequential effects is based on the leaky, competing accumulator choice model of Usher and McClelland (2001). These models share the common property that they supplement the accumulator model with additional traces that track properties of the sequence and offset the decision dynamics from one trial to the next. Cho et al. (2002) demonstrate that individuals maintain multiple biases relevant to different properties of the sequence. These biases, which are implemented via simple decaying memory traces, preferentially contribute to the accumulation rate of each decision unit in a way that depends in part on the identity of the previous trial. Jones et al. (2002) add conflict monitoring to the model of Cho et al. to capture sequential effects in RT that can be attributed to higher-level issues of cognitive control (e.g., speedup and slowdown as the properties of the task change). The model is identical to Cho et al. except that an additional strategic priming bias is added to the accumulation equation. Strategic priming is defined as an exponentially decaying trace of conflict between competing decision units over past trials. Gao et al. (2009) extend this model further to account for modulations due to varied RSI, but the mechanisms of their model are still rooted in simple decaying memory traces.

Two models that are similar to the accumulator model of Usher and McClelland (2001) are the random walk decision model of Laming (1968) and the diffusion model of Ratcliff et al. (1999). Sequential effects are addressed in the context of both of these models though neither makes strong claims about the specific mechanisms that drive the sequential effects. Because no strong claim is made about sequential effects, it is a not completely fair to classify them under the neural inertia category. However, the most natural way to give a mechanistic explanation of sequential effects in these models is to incorporate a bias variable that affects the starting point of the random walk or diffusion process and exhibits neural inertia.

Laming's (1968) random walk model for two-choice (and multi-choice) tasks proposes a decision trace that follows a random walk through the decision space until it reaches one of the absorbing boundaries at which point the decision is made. The increments of the random walk are driven by a segmented stream of input values independently sampled from a distribution that is determined by the stimulus present. On average the random walk gravitates toward the correct response, but the variation in input values can slow the decision process or even lead to an incorrect response. The model hypothesizes that the appropriate starting point for the random walk is dependent upon the relative proportion of the two stimuli. However, Laming suggests that sequential effects may arise because subjects "sub-optimally" estimate the response probabilities from the recent sequence. Such an estimate could easily be derived from a decaying memory trace of previous trials and could be used to produce variations in the starting point of the random walk.

Ratcliff (1999) proposes a diffusion model for decision processing that is in essence the continuous-time counterpart to Laming's (1968) random walk model. Ratcliff demonstrates that variability in drift rates and response boundaries are required for the model to produce sequential effects. However a theoretical explanation for what may cause this variability is not given, though again, as suggested above, this variability could result from residual activity.

Though the choice RT domain has been one of the most active for studying sequential effects, the concept of neural inertia has served to explain sequential effects in many other domains as well. For example in the judgment literature, sequential effects such as response assimilation have been explained as a bias driven by an implicit memory of recent events. In their judgment model for absolute identification, Petrov and Anderson (2005) maintain a decaying memory trace for each anchor or response that represents the degree of activation. The probability of selecting a particular anchor on the next trial increases with the strength of its activation trace. Brown et al. (2008) capture sequential effects in their model of absolute identification via a contrast effect resulting from residual activation in the perceptual scale and a decisional bias towards recent stimuli resulting from residual activation present in the decision units of a ballistic accumulator. In the visual search domain, Maljkovik and Nakayama (1994, 1996) propose that RT is affected by priming of the attention-driving feature (color or spatial frequency) and the target position implemented in a decaying short-term memory trace.

Incremental Learning

Under the neural inertia explanation, the mechanisms responsible for sequential effects are hardwired. Though this is a simple, efficient solution to improving performance when successive events tend to be correlated, it offers no sensitivity to actual performance. An alternative perspective that suggests more engagement with the individual's trial-to-trial performance portrays sequential effects as a consequence of incremental learning processes. As individuals engage in a stream of events, they incrementally adapt behavior in such a way as to reduce future errors or increase future rewards. From this perspective, sequential effects are dependent the sequence of stimuli and responses as in the neural inertia explanation. However, the effects also depend on the appropriateness of past responses. The key property of these models is that the process of long-term learning also indirectly produces short-term sequential effects. Figure 3 displays a simple network architecture in which incremental learning might be used to update weights following each trial according to the error difference between the network output on a trial and the correct response.

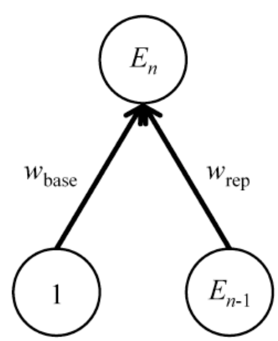

Figure 3. A simple model in which the predicted event at time n, E_n, depends on the previous event and a constant cue. Following the trial, an error signal is obtained by computing the difference between the predicted event outcome and the true outcome. This error signal is used to update weights in an incremental fashion.

The connection between sequential effects and theories of incremental learning has long been recognized. Statistical learning theory (Estes, 1950; Estes & Burke, 1953) hypothesizes a response strategy where response probabilities are updated on a trial-to-trial basis according to the reinforcing events. Because response probabilities fluctuate with each trial presentation, this model predicts simple sequential effects in event prediction paradigms (Estes, 1957; Estes & Straughan, 1954) and in probabilistic discrimination learning paradigms (Estes & Burke, 1957). Taking the analysis a step further, Anderson (1959) derives relationships between parameters of similar incremental learning models and observable sequential effects demonstrating the importance of sequential effects in understanding and testing general models of learning.

Incremental learning is also critical to classical conditioning theory in which the association between a stimulus and an outcome strengthens in a way proportional to how unexpected the outcome was given the stimulus (Rescorla & Wagner, 1972). In connectionist models, this corresponds to a delta-learning rule or error-correction learning and is employed in several models for categorization (e.g., Gluck and Bower, 1988; Kruschke, 1992; Nosofsky, Kruschke, and McKinley, 1992). These models focus on explaining the time course of learning during an experimental session or even across sessions. However, because error-driven learning occurs on every trial, these models naturally exhibit behavior that varies as a function of the sequence of recent stimuli. Jones and Sieck (2003) demonstrate that this class of models accurately capture the positive recency effects found in typical categorization experiments but demonstrate that the models fail in producing the effects found in experiments that manipulate the correlation between outcomes on successive trials. By expanding the ALCOVE model of Kruschke (1992) to include an additional simple statistic that encodes aggregate information about the recent sequence of trials, Jones and Sieck (2003) account for the extra variation in the data that results from sequential autocorrelation.

In the motor control literature, movement error has been accounted for using an autoregressive model that includes past errors. Scheidt et al. (2001) obtain better fits to human movement error in a reaching task with artificially imposed forces by accounting for sequential effects in the data via a regression model that includes current force, previous force, and previous error. Similarly, Wong and Shelhamer (2011) account for the effect of the saccade errors in a repetitive eye-movement task using a simple autoregressive moving average (ARMA) model, though they give evidence for a longer-range learning process as well. Unlike in the connectionist models where weights are adjusted in accordance to the strength of the error signal, in these models, the next behavior depends directly on the previous error. Still these models can be viewed as a coarse form of incremental learning where behavior is continually adjusted according to recent performance, not just recent experiences.

The concept of incremental learning is also embodied in reinforcement learning (Sutton & Barto, 1998) where internal state/action pairs are updated after each experience according to the reward associated with the experience. Because decisions are biased towards recently rewarded actions, these models capture the types of sequential effects in which individuals tend to repeat previously successful actions. For example, reinforcement learning models have been successfully explained sequential effects in shot choice by basketball players (Neiman & Loewenstein, 2012), and attention allocation in a visual cueing paradigm (Chukoskie et al., 2012).

Prediction in a Dynamic Environment

Theories for incremental learning have been popular in part because they represent a simple way for the brain to obtain approximate solutions to complex optimization problems under the constraints of limited memory. Often, models based on incremental learning are proposed for these reasons and the sequential effects they produce are viewed as an inconsequential byproduct of the machinery. Similarly, under the neural inertia account, sequential effects are cast as a secondary byproduct of cognitive mechanisms. Generally these mechanisms are portrayed as suboptimal because they can lead to irrational behavior in certain settings. The classic example is a two-alternative forced-choice (2AFC) task where there is no correlation between successive stimuli and the stimuli occur with equal frequency. Subjects who explicitly know that the stimuli are completely random still exhibit a bias toward one stimulus or the other depending on the recent history of trials (e.g., Cho et al., 2002). This has led many researchers to view sequential effects as a cognitive deficiency. However, if successive events tend to be positively correlated—as seems to be the case in the majority of real-world situations—sequential effects could reflect an intelligent means for adjusting behavior.

Rather than characterizing sequential effects as suboptimal, idiosyncratic behavior, an alternative perspective that has recently gained traction in the literature explains sequential effects as the reflection of a cognitive system designed to adapt to a dynamic environment (i.e., one in which the properties or statistics regularly change). The models that embody this perspective typically begin with basic assumptions about how the statistics of the environment change and then demonstrate that sequential effects result from optimal or approximately optimal behavior given these assumptions. While the neural inertia and incremental learning explanations fall into the implementation or algorithmic levels of Marr's (1982) analysis, the present explanation addresses sequential effects from a computational level of analysis. Here we are less interested in the mechanisms that produce sequential effects and instead seek to understand their critical function or purpose. Figure 4 gives an simple depiction of adaptation to the environment in a 2AFC task.

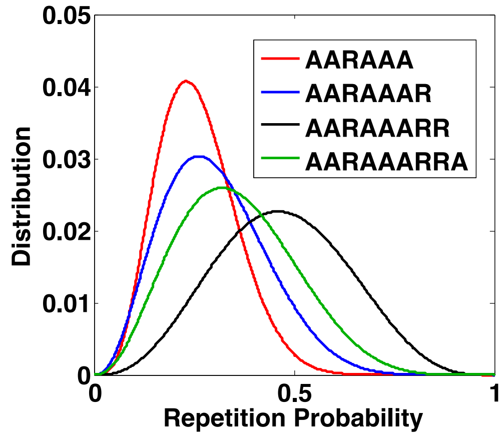

Figure 4. An example of adaptation to the statistics of the environment. Here an individual is assumed to maintain a probability distribution over the probability that the next trial will be a repetition in a 2AFC task. After the sequence AARAAA, there is more density in the lower repetition probabilities. After two Rs, the densities shift toward greater values, but with the final A, the distribution shifts back slightly. Statistics are continually updated as trials occur and a bias is placed on more recent trials such that the model will be flexible to changes in the environmental statistics.

The perspective that behavior reveals underlying assumptions of environmental nonstationarity has been developed in the literature through the years. In considering the seemingly suboptimal behavior of probability matching, where subjects roughly match the frequency of their responses to the outcome probabilities instead of always predicting the most probable response, Flood (1954) suggested that this behavior is appropriate if the subjects expect outcome probabilities to change over time. Nissen and Bullemer (1987) propose that implicit learning of environmental statistical regularities allows for more accurate prediction of events in the near future leading to faster, potentially more accurate behavior. Further, Anderson and Schooler (1991) provided an ecological justification for the functional property of memory decay by comparing it to the empirical need probabilities for information in a variety of real-world domains. This sort of normative argument is compelling because it justifies aspects of cognition according to the contexts in which they are meant to function. As Jones and Sieck (2003) pointed out, the domain of sequential effects in decision-making was in need of such normatively inspired theories.

Recently there has been a proliferation of models that explicitly take this normative approach. Yu and Cohen (2009) explain the classic RT pattern in 2AFC with their Dynamic Belief Model (DBM). The fundamental assumption in the DBM is that the repetition probability in a sequence of trials follows changepoint dynamics—i.e., the probability of the current stimulus/response being the same as the previous is fixed for a period of time until it is randomly resampled. A rational agent operating under these assumptions produces sequential effects that closely match human data. Wilder et al. (2010) extend the DBM exhibiting a better fit to experimental data under the assumption that sequences result from the combination of the repetition probability component of DBM and a baserate stimulus probability also subject to changepoint dynamics. Gokaydin et al. (2011) demonstrate that only baserate probabilities need to be tracked to account for subject RTs in a 3AFC. Though their model is not specifically based on normative principles about the change dynamics, the model predictions are consistent with what would result from the baserate component in Wilder et al. (2010) if it were extended to a three choice task. Though the assumption of abrupt changepoint dynamics has been adopted in these models, it is equally feasible that the environmental statistics change gradually according to random walk dynamics instead (Jones et al., 2011). In many cases, these two different assumptions predict the same pattern of sequential effects though there should be testable differences. In several multiple cue probability learning experiments where the cue-criterion is nonstationarity, Speekenbrink and Shanks (2010) demonstrate that subjects adapt predictions to both gradually changing dynamics and abrupt changepoint dynamics. The authors find some success in modeling the data using a Bayesian linear filter model which essentially uses a Kalman filter to track a random walk in cue criteria, but the model does not conclusively perform better than other existing models suitable for the domain.

In addition to estimating simple sequence statistics like the baserate and repetition probability, individuals have been shown to change behavior according to more abstract properties of the sequence. For example, in difficulty manipulations that intermix easy tasks with similar hard tasks, subjects' RT and accuracy vary depending on the difficulty of recent trials (Taylor & Lupker, 2001). Mozer et al. (2007) propose that these sequential effects arise because individuals estimate task difficulty from recent experience. The model modulates performance on the current trial by taking a weighted average between a current accuracy trace and a historical accuracy trace in a hidden Markov model that is a generalization of a basic accumulation model. Jones et al. (2009) explain the same data with a rationally motivated model founded on the assumption that subjects use the recent history to estimate the drift rate for the correct response—i.e., task difficulty—in a multi-response diffusion model. In a study of difficulty manipulation in a categorization task, Brown and Steyvers (2005) use a changepoint detection model to capture the shift in subject behavior that occurs when there is a shift in the discriminability of the two categories (i.e., how much overlap there is between the continuous stimuli from each category). Steyvers and Brown (2006) take the changepoint detection hypothesis a step further in an experiment where subjects predicted the next item in a sequence generated using changepoint dynamics. The authors show that subjects' behavior corresponded closely to a Bayesian optimal model performing changepoint inference, though it was necessary to parameterize the model sub-optimally to reproduce individual subject's tendency to over or under react to real changes. Interestingly, the ability for tracking environmental changepoints seems to be fundamental for most animals—even rats have been shown to perform close to Bayesian optimal detection of changes in reward rates (Gallistel et al., 2001).

The difficulty manipulations described above require the individual to monitor the properties of the task. Subjects are not told when the difficulty changes and may not even be explicitly aware of the changes, but their behavior reveals that they are tracking task difficulty. In a similar paradigm that studies cognitive control, subjects are given several different tasks to perform and are explicitly aware of task switches. Even still, sequential effects are commonly found in trials following task switches because subjects are not able to change their behavior instantaneously and confuse the previously relevant strategy with the appropriate strategy for the current task. Reynolds and Mozer (2009) take a similar normative approach toward explaining this sort of cognitive control. Specifically, they obtain close fits to human data in a task-switching paradigm by modeling control as a dynamical inference process over the current task. Essentially they characterize cognitive control as an adaptive process that anticipates and responds to changes in the constraints of the environment.

The current explanation of sequential effects as adaptation to changing environmental statistics is not completely at odds with the previous two explanations. In fact, neural inertia and incremental learning can be viewed as implementation- and algorithmic-level solutions to the computational level problem of adapting to a dynamic environment. Yu and Cohen (2009) demonstrate that the DBM is well approximated by an exponentially decaying sum of past trials. This exponentially decaying sum could be implemented in the brain via exponentially decaying residual activation. Incremental learning methods also produce a similar exponentially decaying dependence on past trials. It may well be that our cognitive system is designed to be flexible to a dynamic environment, but lacking the ability to represent the complete statistics of the environment and compute the correct behavior given those statistics, the brain resorts to simpler solutions to produce behavior that is still flexible and adaptive. Nonetheless, the computational perspective that sequential effects reflect an attempt to optimally adapt to changing statistics offers important guidance in understanding why sequential effects are so ubiquitous in behavior.

Sensitivity Adjustment

There is a class of models, mostly in the judgment domain, that address sequential effects from a different angle. These models share the common property that sequential effects result from continual adjustments made to the sensitivity of perceptual and decisional processing. Though sensitivity adjustment can be viewed as a form of adaptation—i.e., the sensitivity is being adjusted in a way that is more appropriate for the type of events likely to occur—this perspective has different implications for how sequential effects manifest.

Perhaps the most familiar example of sensitivity adjustment is found in visual system's high sensitivity to changes in intensity across a broad range of luminance levels. To achieve this sensitivity, many adaptive adjustments occur within the visual pathway as the illumination conditions of the environment change (Ver Hoeve, 2007). In the judgment domain, a classic theory of sensitivity adjustment is the range-frequency model of Parducci (1965). The model assumes two principles that combine to direct how the range of categories is used. The *range principle* assumes category boundaries that depend only on the extreme stimulus values and the number of categories. The *frequency principle* assumes that each category is used for a fixed proportion of the judgments. Whereas the range principle is independent of the stimulus distribution (apart from the extreme values), the frequency principle hypothesizes that the category boundaries shift according to the stimulus distribution. Though sequential effects are not the focus of the model, the frequency principle naturally implies the presence of sequential effects due to changes in sensitivity across the range under the assumption that the boundaries depend more strongly on the recent stimulus distribution because of memory constraints. Figure 5 provides a simplified example how the sensitivity to three different categories might shift after several trials in a categorization task.

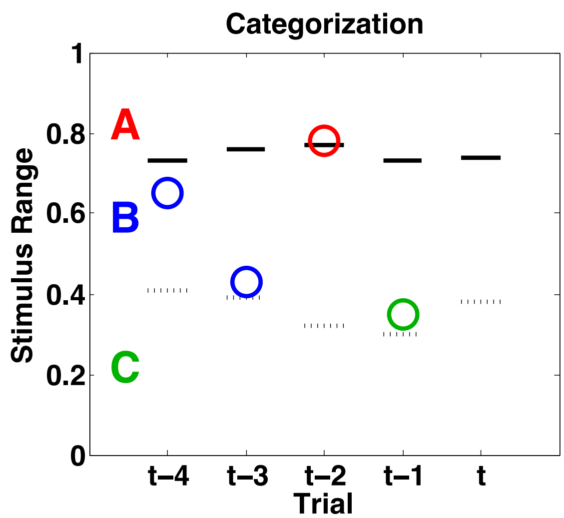

Figure 5. Shifts in category boundaries following a sequence of stimuli. After two presentations of category B, the sensitivity or preference for category B increases. However, with presentations of an A and a C on the subsequent trials, the sensitivity to B decreases again.

From an adaptation perspective, as suggested by Treisman and Williams (1984), the frequency principle serves to maximize the information transmitted by category responses given the current environmental statistics. Similarly, Wainwright (1999) proposes that adaptation in the visual system is driven by the goal to maximize information transmission. At a computational level of analysis, sequential effects that result from range adjustment can be understood as a strategy designed to improve the efficiency of perceptual and decisional processing. Maximizing the information transmitted is one such way in which this efficiency can be improved. Another way to improve efficiency of processing is to incorporate sensitivities to statistics of which types of events are more likely.

Treisman and Williams (1984) build both of these properties into their model for binary categorization. They capture short-term shifts in category boundaries using two competing components—a *tracking* system and a *stabilizing* system—the second of which closely parallels the frequency principle. The tracking system moves boundaries such that recently used categories correspond to a larger range of stimuli. The authors motivate this strategy with the observation that the external world changes: if an object has recently been detected then it is more likely to be detected in the near future and thus the boundary for detection should be lowered to be less restrictive. The fault in the tracking system is that following a series of positive detections, the boundary could be set so low that all discrimination is lost. The stabilizing system counteracts this force by shifting the boundaries such that each category is used with roughly equal probability, i.e., following the frequency principle. In contrast to Parducci (1965), both components of the Treisman and Williams (1984) model adjust sensitivity in response to environmental properties. Advocating the adaptive properties of sensitivity adjustment, Treisman and Williams make the strong claim that "sequential dependencies arise from and reflect the operation of a system that attempts to place criteria at those positions that are optimal at any given moment and to keep them there, and that this same system is at work in all the tasks in which dependencies have been observed." (Treisman & Williams, 1984, p. 93)

Instead of categorizing a stimulus according to a set of boundaries, judgments can be modeled via a set of prototypes to which the stimulus is compared (e.g., Petrov & Anderson, 2005; Smith & Minda, 1998). A shift in the prototype locations can be mapped to an equivalent shift in criteria. However, when feedback is given, prototype locations are defined completely by the preceding stimulus/feedback pairs and thus shifts of the type produced by the frequency principle cannot be produced unless there is a concomitant shift in the feedback distribution. Nonetheless, prototype models have successfully accounted for sequential effects in a variety of identification and categorization experiments. Petrov and Anderson (2005) compute a selection probability for each category as a function of the similarity between the stimulus and the prototype (or anchor as they refer to it) and a category bias that depends on recent usage of the category. The selected anchor shifts after every trial to incorporate the new stimulus value in a way that discounts the previous anchor location exponentially. In this way, the sensitivity of decisional processing is continuously being adjusted to fit the stimulus/response mapping appropriate for the environment. In the absence of feedback, the model has the additional benefit that it adjusts the prototypes to produce balanced category responses under skewed stimulus/response distributions in a way consistent with the Parducci's frequency principle.

One criticism of criterion and prototype models is that they require a fairly large working memory when the set of potential responses or categories is large. Relative judgment theories present a popular alternative that avoids these memory constraints. The basic premise of relative judgment models is that people use the stimulus/response/feedback of the previous trial as a basis for assessing the current stimulus. The sensitivity of processing is continually being adjusted according to the previous stimulus and how it compares to the current stimulus. In some cases, relative judgment strategies also propose trial-to-trial adjustments in the scaling parameter that map a psychophysical difference between stimuli onto the decisional range. In these models, sequential effects generally occur because judgments are based on the memory or perception of the previous stimulus which is often inaccurate because of contamination due to stimuli farther in the past.

Though originally proposed to explain behavior in a magnitude estimation task, where continuous response values are assigned to the stimuli, relative judgment theories are equally applicable to behavior in absolute identification and categorization tasks. Luce and Green (1974) proposed that responses are chosen such that the ratio of the current response to the previous response is proportional to the ratio of the internal representations of the current stimulus and an unfaithful representation of the previous stimulus. The key assumption in the model that produces sequential effects is that the representation of the stimulus from the previous trial is not equal to the stimulus value used to generate the response on the previous trial—i.e., the act of responding to a stimulus changes the representation of the stimulus. DeCarlo and Cross (1990) generalize the response ratio hypothesis by modeling the response as a transformation of the stimulus sensation that depends on a weighted combination of the ratio of the previous stimulus and response and the ratio of a fixed reference stimulus and response. Responses are also affected by an autocorrelated error process. Furthermore, the model assumes that the perception of a stimulus is affected by the context of previous stimuli in such a way that can produce response assimilation or contrast effects depending on parameterization.

Stewart, Chater, and Brown (2005) present a relative judgment model for absolute identification that pushes the relative judgment concept further by assuming that the fundamental currency for perception and decision is the stimulus difference rather than just the pure stimulus. From this perspective, all processing is inherently relative because it depends on the relationship between the previous stimulus and the current stimulus. In their model, this difference probabilistically maps to a response above or below the previous feedback depending on the sign of the difference. The mapping depends on a fixed scaling parameter and a fixed variance noise signal that is distributed over the potential responses. The model also assumes that the stimulus difference on the current trial is confused with or contaminated by the differences on previous trials. This is similar to the context that affects stimulus perception in DeCarlo and Cross (1990), and is primarily responsible for the sequential effects produced by the model because it leads to a contamination in the range of processing.

In addition to shifting the range of processing, relative judgment strategies also impose strong constraints on the range that is relevant for the current trial (e.g., if the current stimulus is perceived as larger than the previous one, only responses greater than the previous feedback will be considered). These constraints generate a shift in sensitivity across the decisional range. Petzold and Haubensak (2001) propose another relative judgment model that places even more constraints on the acceptable response range. In their model, the current stimulus is judged relative to the two previous stimuli and the extreme ends of the range which are assumed to be stored in memory. The appropriate range of potential responses is bounded above and below the current stimulus by the ends of the spectrum or the recent stimuli depending on how the current stimulus compares to the previous two. For example, if the bounds of the range are 1 and 10, the previous two stimuli were 5 and 7, and the current stimulus is perceived smaller than both the previous two, then the range of responses will be 1 to 5. Once the sub-range is determined, the response is selected based on the ratio of the difference between the stimulus and an endpoint and the difference between the sub-range endpoints. The model produces sequential effects because the response depends both on the similarity between the current stimulus and the previous stimuli and the relative location of the stimuli.

The aforementioned models place a greater emphasis on explaining sequential effects that result from adjustments made to sensitivities in decisional processing. However, adjustments made to the perceptual sensitivity can also result in sequential effects. In some cases, it is difficult to separate perceptual from decisional effects because stimuli and responses are confounded in most experiments. Jesteadt, Luce, and Green (1977) further develop the response ratio hypothesis of Luce and Green (1974) and explore the weights of the previous stimulus and response in a regression analysis. They obtain a positive coefficient for the previous response, suggesting assimilation to the previous response, and a negative coefficient for the previous stimulus, suggesting stimulus contrast. This result opens the possibility that perceptual and decisional effects may compete with each other. However, DeCarlo and Cross (1990) have shown that caution must be taken when analyzing the coefficients in a regression equation because different theoretical perspectives lead to different interpretations of the parameters. Nonetheless, perceptual contrast appears to be present in many judgment tasks. Petzold (1981) base their model for categorization on perceptual contrast and a guessing strategy that produces assimilation to the previous response. The model proposes that the trace of the previous stimulus serves as an internal standard for judgment. Helson (1964) has suggested that encoding stimuli using such a relative strategy may serve to expand the representational range of a finite neural system and that perceptual contrast effects may be a consequence of this relative strategy.

Recently, stronger evidence has been provided for perceptual contrast in categorization and identification paradigms via experimental designs that avoid the confound between stimuli and responses (Jones, 2009; Jones, Love, & Maddox, 2006; Jones & Sieck, 2003). By imposing overlapping probabilistic categories or identification labels, it is possible to isolate the effect of the previous stimulus on behavior by considering all trials where the previous feedback label is inconsistent with the previous stimulus. Jones et al. (2006) propose a mathematical model for quantifying the perceptual contrast in experimental data. Though the model does not take a strong theoretical stance, the perceptual component is consistent with the relative judgment model of DeCarlo and Cross (1990) in which the perceptual range used to assess the current stimulus is skewed by the context of stimuli that were recently presented.

The observation that sequential effects result from adjustments in both perceptual sensitivity and decisional sensitivity suggests a greater level of complexity in the experimental data and implies that greater care must be taken in assessing models of sequential effects. Specifically, models that attempt to capture the aggregate sequential effects present in the data may be misguided because they fail to recognize the possibility that the effects are produced by multiple mechanisms operating in concert. Recognizing this division also opens the door for interpreting sequential effects via multiple high-level explanations that play out in different stages of the processing. In fact, this is the stance presented in Jones et al. (2006) and Jones (2009) in which sequential effects are explained as arising from perceptual sensitivity adjustment combined with the notion of decisional generalization. This notion of generalization is the final high-level explanation we present for sequential effects.

Generalization

In classical conditioning, stimulus generalization refers to the tendency for the conditioned stimulus to evoke similar responses after the response has been conditioned and for similar stimuli to evoke similar responses. The degree to which knowledge about one stimulus will generalize to another stimulus has been shown to depend on the similarity between them, with the probability of generalization decreasing exponentially with psychological distance (Shepard, 1957, 1987). In judgment tasks, where subjects categorize, identify, or make estimates about a stimulus and are conditioned through feedback, stimulus generalization can have a significant impact on behavior. Typically, stimulus/response associations are assumed to be relatively constant following a learning period. However, if generalization has a greater dependence on recently conditioned stimulus/response pairs, we would observe sequential effects in the data. Typically, generalization produces an assimilation effect where the response for the current stimulus is biased towards recent responses to similar stimuli (though when recent stimuli are highly dissimilar to the current stimulus, there is some evidence that the effect is contrastive; Jones et al., 2006). Figure 6 portrays a simple generalization process in action.

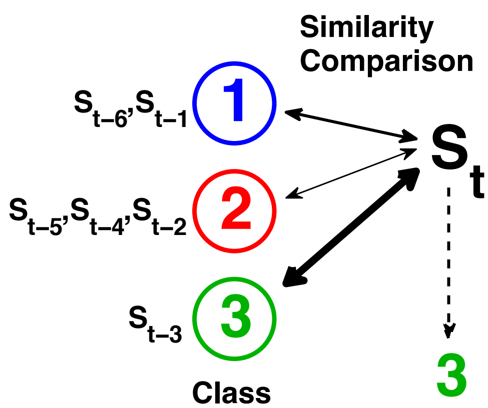

Figure 6. Generalization to past trials. Here previous trials are stored in memory and used as exemplars for three classes. A new stimulus, S_t, is compared to the exemplars and the class label chosen corresponds to the exemplar that best generalizes to the current stimulus.

The concept of generalization is strongly embodied in exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1986) in which decisions result from direct comparisons between the current stimulus and previous stimuli. In their simplest form, exemplar models do not produce sequential effects because stimuli that occurred far in the past have an equal affect on behavior as recent stimuli. However, when the models are modified so that more recent exemplars have a greater likelihood of being recalled or a stronger memory trace, the current response is biased toward recent responses (Nosofsky, Kruschke, & McKinley, 1992; Nosofsky & Palmeri, 1997). Prototype models (e.g., Petrov & Anderson, 2005; Smith & Minda, 1998) are also consistent with the concept of generalization if the role of prototypes matches the role of exemplars above (i.e., judgments are based on the degree to which the different prototypes generalize to the current stimulus). For example, in Petrov and Anderson's (2005) absolute identification model, an anchor (prototype) selection process essentially measures how well each anchor location generalizes to the current stimulus and probabilistically chooses a response. In fact, the prototypes themselves represent a slightly different form of generalization in which knowledge from the collection of past experiences is generalized into a single representative for the class. As new experiences are encountered, they are continually integrated into the anchor estimates in a way that places more weight on recent experiences and consequently results in sequential dependencies similar to those in the exemplar models.

Recent models of category learning have taken an even stronger stance on the relationship between generalization and sequential dependencies (Jones, 2009; Jones et al., 2006; Jones & Sieck, 2003). Jones et al. (2006) demonstrate the dependence of decisional recency on similarity and posit that the effect is a direct byproduct of generalization. Furthermore, they present a mathematical model for quantifying the strength of generalization by measuring the decisional recency effect. Jones (2009) expands these ideas in a model for absolute identification based on the assertion that decisional processes are modulated by similarity-based generalization. In this model, generalization is implemented using reinforcement learning—an incremental learning strategy discussed above. In fact, reinforcement learning is deeply tied to generalization throughout the conditioning literature. Here the computational-level explanation for the sequential effects is that generalization of recent experience affects behavior. However, this perspective is compatible with an algorithmic-level explanation that posits incremental learning processes.

Generalizing in a way that is biased towards recent experience can also be viewed as an adaptive strategy. Jones et al. (2006) suggest that this sort of generalization is well suited for a changing environment. Similarly, Petrov and Anderson (2005) hint that the continual updating of anchor locations helps maintain their utility as the environmental statistics shift

Discussion

We have argued that the broad, diverse literature on sequential effects can be unified under a small set of theoretical accounts for their existence. Rather than viewing these effects as idiosyncratic behavioral modulations unique to each different domain of study, sequential effects can be thought of as reflecting a few high-level computational goals of the cognitive system. This integration brings coherence to sequential effects phenomena and serves researchers in interpreting sequential effects in a way that is consistent and unified across the myriad domains in which they occur.

In our first computational-level explanation, we understand sequential effects by taking the perspective that individuals are trying to form expectations about what stimulus will occur next. Further, we hypothesize that the individuals assume that the environment is changing and thus base these expectations on the sequence of previous stimuli with greater weight placed on recent stimuli. The models that embody this perspective begin with assumptions about the environment and then show that behavior is optimal or near optimal given these assumptions. This perspective naturally extends to the real world where individuals are continually inundated with a barrage of sensations and are required to produce appropriate actions. By preparing ourselves for the situations that are likely to occur next, we can improve the efficiency of our behavior. If expectations are formed with greater emphasis placed on recent experience, our behavior will be more adaptive to the frequent changes in the statistics that govern our environment.

The second computational-level explanation for sequential effects characterizes them as resulting from adjustments made to the sensitivity of perceptual and decisional processing. Rather than forming direct expectations about what events might occur next, here individuals change how they process stimuli according to the stimuli and responses recently experienced. These adjustments can be interpreted as way to improve the efficiency of stimulus processing and response formulation, for example, maximizing the information transmitted by responses, expanding the representational range, or leveraging knowledge about the likelihood of different events.

In our third computational-level explanation, we cast sequential effects as a consequence of generalization. Generalization refers to the process of applying previous knowledge or experience to the current situation. For example, in a judgment study, one may generalize their knowledge with past stimulus/response pairs to form a judgment for the current stimulus. If individuals have a stronger preference to generalize using more recent experiences, sequential effects will appear in the behavioral data. Typically generalization is dependent on the similarity between the current stimulus and the generalizing stimulus. This same dependence on similarity is found in some sequential effects studies, further supporting the presence of generalization.

In addition to the three computational-level explanations for sequential effects, we observed two lower-level explanations present in a broad collection of theoretical accounts. In many models, sequential effects are explained as the result of neural inertia in the cognitive system. The basic priming that results from these mechanisms can explain many types of sequential effects. However, because it is a simple implementation-level account, this explanation fails to provide strong claims about the underlying purpose of sequential effects. Similarly, sequential effects have been explained at an algorithmic level as arising from incremental learning processes in the brain. Again sequential effects are accounted for mechanistically, but there is no strong theoretical stance on why they occur. There is still value in these lower-level explanations for sequential effects because in most cases they are consistent with the three primary computational-level explanations we propose. Though it is important to understand how sequential effects are produced by actual mechanisms in the brain and algorithmic processes, we believe there is significant utility in understanding and classifying sequential effects at a computational level because this level strongly addresses the questions of why sequential effects occur and why they are important.

Other salient distinctions can be made between models of sequential effects that are interesting and informative. However, we have avoided addressing these up to this point so as to not muddle our categorization. For example, many models differ in the duration of memory of past experience and the persistence of the sequential effects. The majority of models posit that sequential effects are very short lived. However, some recent models propose that implicit memory of the past can last over a longer duration or across many intervening events thus resulting in temporally extended sequential effects. Exemplar models are the extreme example of long-lasting memories—i.e., in theory all events are remembered forever. Though, as mentioned earlier, to produce sequential effects, these models must place greater weight on recent trials. Typically, the weights decay exponentially and rapidly go to zero resulting in only short-lasting sequential effects, but a weighting that falls off more slowly would reasonably produce longer-lasting effects. Petrov and Anderson (2005) capture sequential effects in their model in part by maintaining activation levels for each anchor that increase the availability of that anchor on future trials. The activation of an anchor roughly exhibits power function decay dependent on the time elapsed since past uses of that anchor. With its characteristic heavy tail, power function decay results in a longer-lasting memory of the sequence of responses or anchors selected. Wong and Shelhamer (2011) similarly propose power function decay over past trials and demonstrate correlations in behavior that extend out to almost 100 intervening trials. Wilder et al. (2012) offer further evidence for longer-lasting sequential effects in a 2AFC task and a motor control task. Building upon theories that assume a dynamic environment, they propose the Hierarchical Dynamic Belief Model (HDBM) that relaxes unnatural assumptions in the DBM and results in an integration of past trials that more closely resembles a power function weighting.

Another distinction between models is the degree of abstraction used for representing trials or events. The simplest representation is the actual trial identity (e.g., a specific digit between 1 and 6 as in Falmagne, 1965; or light 2 and light 3 in Remington, 1969). Alternatively, trials can be represented in how they relate to the previous trial. This is most sensible in 2AFC tasks where each trial is either a repetition or an alternation (e.g., Hale, 1969, studies the effect of runs of repetitions and alternations). Depending on which abstraction is used in the analysis, the pattern of sequential effects can appear very different. Cho et al. (2002) demonstrate the value of considering multiple abstractions by modeling sequential effects via the combination of a mechanism that tracks the actual trial identities and a mechanism that tracks the repetition/alternation sequence. Even more complex abstractions can be responsible for sequential effects as in models that capture dynamic changes in task difficulty (e.g., Jones et al., 2009; Mozer et al., 2007). There is less evidence of representational abstraction influencing the type of effects observed in judgment studies. However, Stewart et al. (2005) propose that the currency used in processing a stimulus is its difference relative to the previous trial rather than its pure psychophysical magnitude. Perhaps there are other levels of abstract representations that influence effects in judgment tasks, but have yet gone unnoticed. For example, how might task difficulty modulations change the sequential effects observed in an identification or categorization task? Brown and Steyvers (2005) suggest that there will be an effect, but they only consider changes in difficulty that occur across blocks of trials, not within blocks.

The difference between the neural inertia explanation and the incremental learning explanation hints at another distinction that can be made between models. Unlike in the neural inertia account where only the sequence of previous trials affects behavior, with incremental learning, modulations in behavior are based on both the sequence of past trials and the individual's performance on those trials. This distinction can be cast as a difference between unsupervised learning and supervised learning strategies. Models that utilize unsupervised learning learn from the environment without feedback. Neural inertia and short-term priming models have this characteristic as do the more complex models that perform Bayesian inference over the environmental statistics assuming nonstationarity. Supervised learning is present in incremental learning models and can be present in generalization models that utilize reinforcement learning (e.g., Jones, 2009). Most models in the sensitivity adjustment explanation category can be viewed as unsupervised learning models (e.g., the tracking and stabilizing systems in Treisman & Williams, 1984, are unsupervised). Nonetheless, it is feasible that adjustments to sensitivity could be made in a supervised manor.

Sequential effects have been attributed to another factor that we have not yet discussed. Specifically, correlations in successive responses can arise from latent variables in the cognitive system that influence behavior but change over longer timescales. For example, in a task where the subject simply presses a button each time a single stimulus appears, it is likely that the subject's attention will slowly drift throughout the experiment resulting in nearby trials to have similar characteristics (i.e., faster when attention is high and slower when attention is low). In the motor control literature, Kording et al. (2007) propose a model that uses a Kalman filter to capture the influence on behavior of multiple latent variables which drift at varying timescales. Similarly, DeCarlo and Cross (1990) incorporate an autocorrelated error process into their model for sequential effects in judgement. The authors explicitly discuss how this error process can capture the effects of slowly drifting latent variables. Gilden et al. (1995) have demonstrated the presence of 1/f noise in many aspects of cognition (i.e., long-range correlations in behavior that result from some property of the internal state). Though this type of response autocorrelation is often classified as a sequential effect, it not consistent with other types of sequential effects because behavior is not modulated by the exact sequence of recent trials. Nonetheless, it is important to identify this response autocorrelation and be aware of how it can change the appearance of other sequential effects.

In this article, we have demonstrated the widespread presence of sequential effects in cognition and the lack of organization that exists in the large collection of models that seek to explain these effects. Our goal has been to introduce a theoretical framework that synthesizes this disparate literature and highlights a few high-level principles that explain sequential effects across all domains of cognition. In the process, we have observed that most sequential effects reflect an adaptive process in the brain. This is most evident in our first computational-level explanation in which sequential effects result from the attempt to form expectations of future events under the assumptions of a dynamic environment. By studying sequential effects through the lens of the models that fall in this category, it is possible to expose what individuals are adapting to and how they go about that adaptation. In the other two computational-level explanations, an adaptive process is suggested, but the claims about the specifics of adaptation are weaker. However, by founding these sorts of models more directly on normative principles regarding adaptation to the environment, it may be possible to learn more about the dynamics of cognitive adaptation. We believe this work will provide a useful framework for understanding sequential effects in all domains and we hope that we will encourage researchers to pay closer attention to sequential effects and give greater consideration to what insight these effects might offer with respect to the role of adaptation in their specific domain.

References

Anderson, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), Studies in mathematical learning theory, Stanford: Stanford University. Press, 1959.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2, 396-408.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. Quarterly Journal of Experimental Psychology, 13, 90-102.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? Journal of Experimental Psychology: General, 129, 177-192.

Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. Psychological Review, 115, 396-425.

Brown, S.D., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 587-599.

Cho, R. et al. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. Cognitive, Affective, & Behavioral Neuroscience, 2 (4), 283-299.

Chukoskie et al., (in preparation)

Chun, M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. Cognitive Psychology, 36, 28-71.

Cuthill, I. C., Kacelnik, A., Krebs, J. R., Haccou, P., & Iwasa Y. (1990). Starlings exploiting patches: The effect of recent experience on foraging decisions. Animal Behaviour, 40, 625-640.

DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. Journal of Experimental Psychology: General, 119, 375-396.

Doshi, A., Tran, C., Wilder, M., Mozer, M. C., & Trivedi, M. M. (2012). Sequential dependencies in driving. Cognitive Science, 36(5).

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychol. Rev.,* 57, 94-107.

Estes, W. K. (1957). Theory of learning with constant, variable, or contingent probabilities of reinforcement. Psychometrika, 22, 113-132.

Estes, W. K., & Burke, C. J. (1953). A theory of stimulus variability in learning. *Psychol. Rev.,* 60, 276-286.

Estes, W. K. and Straughan, J. H. (1954). Analysis of a verbal conditioning situation in terms of statistical learning theory. J. exp. Psychol., 47, 225-234

Falmagne, J. C. (1965). Stochastic models for choice reaction time with applications to experimental results. Journal of Mathematical Psychology, 12, 77-124.

Flood, M. M. (1954). Environmental non-stationarity in a sequential decision-making experiment. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), Decision processes (pp. 287- 299). New York: Wiley.

Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. Journal of General Psychology, 113, 351-357.

Gallistel, C., Mark, T., King, A., & Latham, P. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. Journal of Experimental Psychology, Animal Behavior Processes, 27 (4), 354-372.

Gao, J., Wong-Lin, K.F., Holmes, P., Simen, P. & Cohen, J. D. (2009). Sequential effects in two-choice reaction time tasks: Decomposition and synthesis of mechanisms. Neural Computation, 21(9), 2407-2436.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. Journal of Experimental Psychology, 46, 373-380.

Gilden, D.L., Thornton, T., & Mallon, M.W. (1995). 1/f noise in human cognition. Science, 267, 1837-1839.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology: General, 117, 227-247.

Gokaydin, D., Ma-Wyatt, A., Navarro, D. J. & Perfors, A. F. (2011). Humans use different statistics for sequence analysis depending on the task. Proceedings of the 33rd Annual Conference of the Cognitive Science Society.

Hale, D. J. (1969). Repetition and probability effects in a serial choice reaction task. Acta Psychologica, 29, 163-171.

Helson, H. (1964). Adaptation-level theory. Oxford, England: Harper & Row.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief adjustment model. Cognitive Psychology, 24, 1-55.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. Journal of Experimental Psychology, 45, 188-196.

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. Journal of Experimental Psychology, 41, 291-297.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. Journal of Experimental Psychology: Human Perception and Performance, 3, 92-104.

Johnson, J., Tellis, G. (2005). Blowing Bubbles: Heuristics and Biases in the Run-Up of Stock Prices. *Journal of the Academy of Marketing Science*, 33, 486-503.

Jones, A. D., Cho, R. Y., Nystrom, L. E., Cohen, J. D., & Braver, T. S. (2002). A computational model of anterior cingulate function in speeded response tasks: Effects of frequency, sequence, and conflict. Cognitive, Affective, & Behavioral Neuroscience, 2, 300-317.

Jones, M. & Sieck, W. (2003). Learning myopia: An adaptive recency effect in category learning. Journal of Experimental Psychology: Learning, Memory, & Cognition, 29, 626-640.

Jones M, Love BC, & Maddox WT (2006). Recency as a window to generalization: Separating decisional and perceptual sequential effects in categorization. J Exp Psychol Learn, 32, 316-332.

Jones et al., 2011 (poster presentation)

Jones, M. (2009). A Reinforcement-and-Generalization Model of Sequential Effects in Identification Learning. Annual Meeting of Cognitive Science Society. pp. 1180-1185.

Kirby, N. H. (1976). Sequential effects in two-choice reaction time: Au- tomatic facilitation or subjective expectancy? Journal of Experimental Psychology: Human Perception & Performance, 2, 567-577.

Kording, K. P., Tenenbaum, J. B., & Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. Nature Neuroscience, 10, 779-786.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review, 99, 22-44.

Link, B. V., Kos, B., Wager, T. D., & Mozer, M. C. (2011). Past experience influences judgment of pain: Prediction of sequential dependencies. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), Proceedings of the 33d Annual Conference of the Cognitive Science Society, 1248-1253. Austin, TX: Cognitive Science Society.

Laming, D. (1968). Information theory of choice reaction times. London: Academic Press.

Luce, R. D., & Green, D. M. (1974). The response ratio hypothesis for magnitude estimation. Journal of Mathematical Psychology, 11, 1-14.

Marr, D. (1982), *Vision: A Computational Approach*, San Francisco, W. H. Freeman.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. Psychological Review, 85, 207-238.

Mozer, M., Kinoshita, S., & Shettel, M. (2007). Sequential dependencies offer insight into cognitive control. In W. Gray (Ed.), Integrated Models of Cognitive Systems, 180-193. Oxford University Press.

Mumma, G. H., & Wilson, S. B. (2006). Procedural debiasing of primacy/anchoring effects in clinical-like judgments. Journal of Clinical Psychology, 51, 841-853.

Maljkovic, V., & Nakayama, K. (1994). Priming of popout: I. Role of features. Memory and Cognition, 22, 657-672.

Maljkovic, V. & Nakayama, K. (1996). Priming of popout: II. Role of position. Perception & Psychophysics, 58, 977-991.

Neiman, T. & Loewenstein, Y. (2011). Reinforcement learning in professional basketball players. Nature Communications, 2, Article number: 569. doi:10.1038/ncomms1580

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology, 19 , 1-32.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39-57.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 211-233.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. Psychological Review, 104, 266-300.

Parducci A (1965). Category judgment: A range-frequency model. Psych Rev, 72, 407-418.

Petrov, A. A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. Psychological Review, 112, 383-416.

Petzold, P. (1981). Distance effects on sequential dependencies in categorical judgments. Journal of Experimental Psychology: Human Perception and Performance, 7, 1371-1385.

Petzold, P. and Haubensak, G. (2001). Higher order sequential effects in psychophysical judgments. Perception & Psychophysics, 63 (6), 969-978.

Ratcliff,R.,VanZandt,T.,& McKoon,G.(1999). Connectionist and diffusion models of reaction time. Psychological Review, 106, 261-300.

Real, L. A. (1991). Animal choice behavior and the evolution of cognitive architecture. Science, 253, 980-986.

Remington, R. J. (1969). Analysis of sequential effects in choice reaction times. Journal of Experimental Psychology, 82, 250-257.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), Classical Conditioning II. 64-99. Appleton-Century-Crofts.

Reynolds, J. & Mozer, M. (2009). Temporal dynamics of cognitive control. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), Advances in Neural Information Processing Systems, 22, 2215-2222. La Jolla, CA: NIPS Foundation.

Scheidt, R. A., Dingwell, J. B., & Mussa-Ivaldi, F. A. (2001). Learning to move amid uncertainty. Journal of Neurophysiology, 86 (2), 971-985.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. Psychometrika, 22, 325-345.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science, 237, 1317-1323.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 1411-1436.

Soetens. E., D eboeck, M ., & Hueting, J. (1984). Automatic aftereffects in two-choice reaction time: A mathematical representation of some concepts. Journal of Experimental Psychology: Human Perception and Performance, 10, 581-598.

Soetens, E. , Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. Journal of Experimental Psychology: Human Perception & Performance, 11, 598-616.

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. Journal of Experimental Psychology: General , 139 , 266-298.

Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. Journal of Mathematical Psychology, 49, 403-409.

Steyvers, M. & Brown, S. (2006). Prediction and change detection. Advances in neural information processing systems, 18, 1281-1289.

Sutton, R. S., Barto, A. G. (1998). Reinforcement Learning: An Introduction. MIT Press.

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27, 117-138.

Treisman, M. C., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. Psychological Review, 91, 68-111.

Usher, M., & McClelland, J. (2001). The time course of perceptual choice: The leaky, competing accumulator model. Psychological Review, 108, 550-592.

Ver Hoeve, J. N., Visual Adaptation (2007). Duane's Foundations of Clinical Ophthalmology, Volume 2, Chapter 16.

Vervaeck, K. R., & Boer, L. C. (1980). Sequential effects in two-choice reaction time: Subjective expectancy and automatic aftereffect at short response–stimulus intervals. Acta Psychologica, 44, 175-190.

Vlaev, I., Chater, N., & Stewart, N. (2007). Financial prospect relativity: Context effects in financial decision-making under risk. *Journal of Behavioral Decision Making, 20,* 273-304.

Wainwright, M.J. (1999). Visual adaptation as optimal information transmission. *Vision Research, 39,* 3960-74.

Ward, L. M. (1973). Magnitude estimations with a variable standard: sequential effects and repeated other properties. Perception & Psychophysics, 13, 193-200.

Ward, L. & Lockheed, G. (1970). Sequential effects and memory in category judgements. Journal of Experimental Psychology, 84 (1), 27-34.

Wilder, M., Jones, M., Ahmed, A., Curran, T., & Mozer, M. (2012). To appear.

Wilder, M., Jones, M., & Mozer, M. C. (2010). Sequential effects reflect parallel learning of multiple environmental regularities. In Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), Advances in Neural Information Processing Systems, 22, 2053-2061. La Jolla, CA: NIPS Foundation.

Wong, A. L. & Shelhamer, M. (2011). Exploring the fundamental dynamics of error-based motor learning using a stationary predictive-saccade task. PLoS ONE, 6 (9): e25225.

Yu, A. & Cohen, J. (2009). Sequential effects: Superstition or rational behavior? In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), Advances in Neural Information Processing Systems, 22, 1873-1880. La Jolla, CA: NIPS Foundation.