Psychon Bull Rev (2015) 22:38–53
DOI 10.3758/s13423-014-0669-3

THEORETICAL REVIEW

Optimal decision making in heterogeneous and biased environments

Rani Moran

Published online: 14 June 2014

Abstract The issue of optimal performance in speeded two-choice tasks has played a substantial role in the development and evaluation of decision making theories. For difficulty-homogeneous environments, the means to achieve optimality are prescribed by the sequential probability ratio test (SPRT), or equivalently, by the drift diffusion model (DDM). Biases in the external environments are easily accommodated into these models by adopting a prior integration bias. However, for difficulty-heterogeneous environments, the issue is more elusive. I show that in such cases, the SPRT and the DDM are no longer equivalent and both are suboptimal. Optimality is achieved by a diffusion-like accumulation of evidence while adjusting the choice thresholds during the time course of a trial. In the second part of the paper, assuming that decisions are made according to the popular DDM, I show that optimal performance in biased environments mandates incorporating a dynamic-bias component (a shift in the drift threshold) in addition to the prior bias (a shift in the starting point) into the model. These conclusions support a conjecture by Hanks, Mazurek, Kiani, Hopp, and Shadlen, (*The Journal of Neuroscience, 31*(17), 6339–6352, 2011) and contradict a recent attempt to refute this conjecture by arguing that optimality is achieved with the aid of prior bias alone (van Ravenzwaaij et al., 2012). The psychological plausibility of such "mathematically optimal" strategies is discussed. The current paper contributes to the ongoing effort to understand optimal behavior in biased and heterogeneous environments and corrects prior conclusions with respect to optimality in such conditions.

Keywords Decision making · Optimal performance · Bias · Sequential sampling · Drift diffusion model · Sequential probability ratio test · Temporally flexible choice thresholds

R. Moran (✉)
School of Psychological Sciences, Tel Aviv University, Ramat Aviv, POB 39040, Tel-Aviv 69978, Israel
e-mail: rani.moran@gmail.com

Having ordered fresh fish rather than game for both you and your sweetheart, the only thing that is needed for a perfect dinner is the appropriate wine! When you are just about to obey the dictum, "White wine accompanies fish, red wine–meat" and order a bottle of Chardonnay, as if to make your life more complicated, the wine waiter begins to bombard you with indispensable information. He informs you that the red Merlot won the gold medal in the last wine festival in Venice, that the white Riesling is especially effective in creating a romantic atmosphere, that the fish you ordered has a supreme aromatic taste, etc. This baffling situation demonstrates the real-life necessity to combine prior general knowledge (white wine is generally more appropriate with fish) with situation-specific novel information.

In the cognitive laboratory, participants face similar, albeit generally less romantic, situations. Consider a perceptual, speeded-choice task in which participants are asked to choose which of two lights flickers with a higher rate. Suppose that participants are informed in advance that on 75 % of the trials the left, rather than the right-side, light corresponds to the correct answer. As the trial onsets, participants should integrate trial-specific perceptual (flickering) evidence with the advance information with respect to the external bias. What is the optimal way to make decisions in such situations?

The standard or ideal of "optimality" has exerted a fundamental and prolonged influence on research, guiding both the development and evaluation of cognitive as well as of normative decision-making models (e.g., in psychology, economics). In addition, optimal problem-solving contributes immensely to the development of algorithms in the computer, neural networks and artificial intelligence sciences. Despite its importance, the issue of optimality suffers from misinterpretations that may lead research efforts astray. The purpose of the current paper was to contribute to the ongoing effort to understand optimal behavior in biased and heterogeneous environments (see below) and to correct prior conclusions with respect to optimality in such conditions.

The current research is grounded in the framework of sequential sampling, which has been playing a pivotal role in the study of optimality. Generally, sequential sampling theories assume that participants sample continuously the (perceptual) stimulus, gathering with each "time frame" a novel piece of evidence. This information is accumulated in favor of each of the response alternatives, until some threshold level is reached. This threshold can correspond either to a relative (in the drift diffusion model; DMM; e.g. Ratcliff, 1978; Ratcliff & Rouder, 2000; Wagenmakers, 2009) or an absolute (in race models; e.g. Brown & Heathcote, 2008; Vickers, 1979; Usher & McClelland, 2001[1]) level of evidence in favor of one alternative or the other. Within this framework, it is assumed that prior knowledge with respect to the correct decision biases the response (Diederich & Busemeyer, 2006; Ratcliff & McKoon, 2008; Mulder et al., 2012).

Before identifying the optimal decision rule, the term should be clearly defined. In the current paper, I focus on two definitions of optimality that have been dominating the decision-making literature. According to the first definition (Wald, 1947; Wald & Wolfowitz, 1948), henceforth "Wald optimality" (WO), a decision-making rule is optimal if it *minimizes* the mean response time (MRT) while achieving a desired level of accuracy. Thus, any other decision rule that achieves the same level of accuracy is necessarily more time-consuming (in the mean). A dual perspective on the WO decision rule is that it *maximizes* accuracy for a given MRT. The second definition of optimality, henceforth "reward rate optimality" (RRO), pertains to maximization of the reward rate (RR). According to this definition, behavior is optimal when the decider maximizes the average reward per time-unit (Gold & Shadlen, 2002). RRO is probably a more ecologically valid concept of optimality than WO in that it describes better the goals and incentives of animals and humans in decision tasks (Balci et al. 2011).

The concept of WO has traditionally been associated with the models that achieve it, the sequential probability ratio test (SPRT; Wald and Wolfowitz, 1948; Laming, 1968) and the drift diffusion model (DDM). Recently, however, it has been shown that alternative models are more efficient than DDM with respect to reward rate. Such models include the urgency gating model (Cisek et al., 2009; Thura et al., 2012) and a simplified variant of a Bayesian model (Deneve, 2012). Nevertheless, these models also are reward-rate suboptimal. Drugowitsch et al. (2012) applied the method of dynamic programming to show that RRO is achieved by a strategy of integrating information according to a diffusion-like mechanism. Critically, however, the RRO strategy mandates adjusting the choice threshold within the course of individual trials, unlike DDM in which the choice threshold is constant within trials. In summary, RRO optimization is attained via combining a diffusion-like integration with a within-trial, temporally flexible stopping rule.

The fact that the two notions of optimality, WO and RRO, are associated with different models may lead to the erroneous conclusion that these are two unrelated concepts of optimality, each subserved by its own optimality achieving model(s). However, such interpretation must be avoided. Indeed, in the first part of the paper, I will show that WO and RRO are intimately related, in fact equivalent, concepts. Consequently, achieving one form of optimality satisfies the other as well (for an appropriate selection of parameters, as explained below).

If WO and RRO are equivalent forms of optimality, then how come they are associated with different models? The key to the answer pertains to the distinction between homogeneous and heterogeneous environments. In the context of the current discussion, an environment is simply a set or a "block" of experimental trials. A block of trials is *homogeneous* or *heterogeneous* (with respect to difficulty) if the difficulty level across trials in the block is constant or variable respectively. In discussing heterogeneous environments, I assume that the difficulty level is random, that is, unbeknownst in advance (i.e., on trial-onset) to the decider. However, the distribution of difficulty levels, across trials, is assumed to be known to the observer. For example, if across all experimental trials the two flickering rates are maintained at constant levels, the block of trial is homogeneous. Otherwise, if the flickering rates are subject to random across-trial variability, it is heterogeneous.

Taking the heterogeneity of the environment into account, the confusing association between the two equivalent forms of optimality, WO and RRO, and the various optimal models is resolved as follows: the model that adjusts the choice threshold within the time course of a trial (Drugowitsch et al., 2012) is generally optimal, i.e., it is optimal for both homogeneous and heterogeneous environments. The precise form of choice-threshold adjustment is identified with the aid of a dynamic programming method[2] and varies as a function of environmental properties (e.g., the distribution of difficulty levels across trials). When the environment is heterogeneous, SPRT and DDM are not equivalent models and furthermore, neither model is optimal. However, when the environment is homogeneous, the dynamic-programming solution requires no threshold adjustment within a trial (Drugowitsch et al.), and thus the model reduces to the standard DDM. Finally, in

[1] In Usher & McClelland's Leaking Competing Accumulators model (LCA; 2001), the evidence in favor of each alternative is taxed by mutual inhibition from the other alternative.

[2] Drugowitsch et al. (2012) allowed for the possibility that integration of information is associated with a temporal cost $c(t)$. Throughout the paper, I ignore such costs i.e., assume that $c(t)=0$.

such cases (homogeneous environments), the SPRT and DDM also are equivalent, and hence, all three models are equivalent and optimal.

In summary, RRO is equivalent to WO and thus both are attained by the same model(s). The dynamic programming approach identifies the generally optimal decision rule in any environment and it reduces to the DDM (and SPRT) in homogeneous environments. The goal of the first part of the paper is to clarify these misleading issues, which have confused researchers over the past few years.

To illustrate this confusion, consider the Bayesian reader, an influential model of word recognition (Norris, 2006; Norris, 2009). Norris has successfully applied this model to account for performance in the lexical decision task, i.e., the task of deciding whether a letter-string is an eligible word or a non-word. In a nutshell, during a lexical decision, the model continuously executes a Bayesian computation of the log likelihood ratio (LR) that a probe letter-string is a word vs. a non-word (see Eq. (5) in Norris, 2009) and commits to a decision as soon as the LR reaches an upper (word) or a lower (non-word) choice threshold. Thus, for the lexical decision task, the model is equivalent to SPRT.

Norris (2006, 2009) argues that the Bayesian reader achieves WO for lexical decisions. It is noteworthy that the notion of optimality plays a vital role in the conceptual framework of the ideal observer (Geisler, 2003) in which the Bayesian reader is grounded. Indeed, according to Norris, optimality is a virtue of the Bayesian reader qua theory of behavior, as it enables the theory to account not only for *how* people behave, i.e., how behavior is produced, but also for *why* they behave as they do (both the how and the why are hallmarks of a good theory). Optimal models provide a straightforward answer to the why question: optimal behavior reflects optimal adaptation to the environment. Many alternative (to the Bayesian reader) suboptimal models, however, are confined to explaining the how.

Most relevantly, lexical decision task consists of a heterogeneous environment (unless a single word and a single nonword are tested across trials). For example, the higher a word's frequency, the easier it becomes to discriminate it from non-words (Norris, 2009). Thus, lexical decision tasks that mix words of varying frequency within an experimental block, such as the tasks studied by Norris, are heterogeneous. Consequentially, the SPRT and thus the Bayesian reader, is a *suboptimal* rather than an optimal decision strategy.

Indeed, as Drugowitsch et al. (2012) show, there are two equivalent ways to describe the information-integration process in their optimal model, either as a diffusion-like integration of evidence or as an online calculation of the belief, that is, the probability that the observer assigns to the correctness of each response-alternative.[3] Additionally, beliefs are equivalently described by the LR ratio between the correctness of both response alternatives (because there is a one-to-one mapping between the probability that option A is correct and the LR of alternative A vs. B). Importantly, both the evidence and the belief formulations require an adjustment of an (evidence or belief) choice threshold during the trial. Returning to the question of whether the Bayesian reader attains optimality in the lexical decision task, I note that likelihood ratios are updated by utilizing the Bayes rule and, as such, Bayesian beliefs (such as the beliefs formed by the Bayesian reader) are instrumental to attaining optimality. Critically, Bayesian formations of beliefs are not a sufficient condition for optimality, because optimality mandates an appropriate temporally flexible stopping rule. The Bayesian reader does not satisfy this requirement as it maintains a rigid response threshold during the trial (as SPRT).

Interestingly, Norris (2009) argues that a DDM model applied to the lexical decision task (Ratcliff, Gomez & Mckoon, 2004) also achieve WO. However, this claim is incorrect as the DDM model is not optimal in heterogeneous environments as well (again, due to the lack of threshold adjustment along the trial). Similar claims about the optimality of both the Bayesian reader and the DDM in heterogeneous lexical decision tasks were made by other researchers as well (Wagenmakers, Ratcliff, Gomez & Mckoon, 2008). It thus seems timely to clarify such confusions and particularly to direct attention to the essential distinction between homogeneous and heterogeneous environments.

In the second part of the paper, I will limit my focus to the DDM (rather than consider all potential decision rules). DDM, arguably the most influential of the sequential sampling models, has been highly successful in accounting for performance in heterogeneous environments in a vast plethora of decision tasks (Ratcliff & McKoon, 2008). Thus, it is instrumental to examine the more modest question of identifying the optimal behavior in heterogeneous environments *given that the DDM is the processing architecture*. This investigation will emphasize *biased* heterogeneous environments, because such environments have been in the center of a recent interesting debate, which will be described shortly. An environment is called biased if the *a priori* probabilities that the two response-alternatives are correct are unequal. I assume

[3] Note that the term evidence refers to the variable that is being integrated by the diffuser in order to make a decision (e.g. perceptual samples). Once the statistical properties of such evidence, i.e. its distribution under both response alternatives is specified, the observer can form his or her belief, i.e. calculate the probability of the correctness of each response alternative (or the likelihood ratio), given the stream of evidence that has been collected (see Appendix B). Both processes, integration of evidence or belief-update are subsumed under the general term 'integrating information'.

that the bias is known to the decision-maker in advance. The question of optimality in biased heterogeneous environments restricted to the DDM reduces to the identification of the optimizing DDM parameters (see below).

With respect to DDM, two mechanisms have been suggested for implementing environmental bias: prior and dynamic bias. *Prior bias* manifest when the observer lowers the criterion level of evidence that suffices for choosing the more, relative to the less, *a priori* likely alternative. *Dynamic bias*, on the other hand, takes effect if each piece of accumulated evidence is slanted to some extent in favor of the more *a priori* likely alternative. I dub the aggregated effect of both prior and dynamic bias on evidence accumulation as *integration bias* (to distinguish it from the external bias in the environment). Whereas prior bias is fixed at a constant level along a trial, dynamic bias builds up the integration bias continuously with the passage of time.

How should bias in decision making be implemented if the decision maker is to behave optimally (in the DDM)? In a recent interesting paper, Hanks et al. (2011) suggested that when the environment is heterogeneous, optimal behavior mandates the accommodation of advance information (of an environmental bias) in the form of a dynamic bias in addition to prior bias. While providing no formal proof for their conjecture, these authors offered a compelling intuition: first, the longer the trial lasts without reaching the decision threshold, the more it becomes likely that the trial is difficult rather than easy. Second, the more difficult the trial, the less informative the perceptual evidence is (because the perceptual signal to noise ratio is lower), relative to the advance information and hence, more emphasis should be put on the latter. Putting these pieces together, the later into the trial, the more the advance knowledge of bias should be amplified. The dynamic bias implements such a strategy, because it increases the integration bias with the passage of time.

However, this intuition has been challenged by van Ravenzwaaij et al. (2012). These authors purported to demonstrate that this conjecture is wrong and that in fact, even when a biased environment is heterogeneous, optimality is achieved by incorporating prior bias alone. In the current paper, I show that in accordance with the Hanks et al. (2011) conjecture, performance is optimized by accommodating both prior and dynamic bias as components of the integration bias. I will show that the analysis of van Ravenzwaaij et al. suffered from a few shortcomings and that consequentially, their dismissal of Hank's conjecture was premature. My reanalysis yields an affirmative support for the Hanks et al. conjecture.[4]

[4] I note from the outset that the van Ravenzwaaij et al. article consists of both a theoretical study of optimal behavior (in both homogeneous and heterogeneous biased environments) and an empirical study of actual behavior. Here, I question only the conclusions with respect to the theoretical analysis of the heterogeneous environments.

Drift diffusion model

As much of the current discussion is rooted in the framework of the DDM, I provide a very brief description of this model (for elaboration the reader is referred to Ratcliff, 1978; Ratcliff and Rouder, 2000; Wagenmakers, 2009). In DDM, a single accumulator is updated continuously (in time) as a stream of noisy perceptual evidence is incoming. The state of the accumulator corresponds to the net balance of evidence in favor of one alternative over the other. Two response thresholds correspond to the two response options and are set at evidence levels 0 and a (a is dubbed the boundary separation and corresponds to the response caution of the participant). As soon as the accumulated amount of evidence reaches either of the response thresholds, the corresponding decision is executed. The threshold-reaching time corresponds to the decision latency. The observed response time includes a component of residual time that corresponds to decision-extrinsic processes, such as the initial perceptual encoding of the stimulus, the motor response execution time, etc. However, it is assumed to be an additive component that is independent of the decision. Therefore, any WO decision rule minimizes simultaneously the decision and the observed RT (for a given accuracy rate).[5]

The dynamics of the DDM proceeds as follows: it is assumed that at time $t = 0$, the starting point of the accumulator is $x(0) = z$, $0 \leq z \leq a$. For $t > 0$, it is assumed that the accumulator dynamics evolves according to the equation $dx(t) = vdt + sdW(t)$, where $W(t)$ represent a Wiener noise process ("an idealized Brownian motion"), s represents the standard deviation of that noise process,[6] and v represent the evidence drift rate.

Interrogation vs. free response tasks In the current paper, I will discuss two types of tasks: interrogation vs. free-RT task. The core difference between these two types lies in the locus of control of the response time. Whereas in interrogation paradigms, the participants have to respond as soon as the experimenter issues a response signal, in the free-RT task participants are free to respond when they "feel ready." Below, I will discuss optimality with respect to both tasks. It turns out that identifying the optimal strategy is simpler for interrogation tasks.

In discussing DDM, I always assume that in free RT tasks, the thresholds are temporally constant, i.e., do not change within the time course of trials. Furthermore, I assume that in interrogation tasks, the observer integrates evidence until a

[5] Note that this residual time components is subsumed in the term t_{res} in the definition of reward rate, see Eq. (1) below.

[6] Throughout the paper I follow the customary convention and fix $s = 0.1$. This practice reflects the assumption that the noise level is identical across all conditions (difficulty levels, in the current case). However, mathematically speaking, this procedure poses an 'over-constraint' on the model (Donkin, Brown & Heathcote, 2009)

response signal is issued at time T,[7] and a decision is made according to the sign of $x(T)$.

Biases in the DDM How are prior and dynamic biases implemented in the DDM? Prior bias is implemented by the parameter z. When $z = {}^{a}/_{2}$, the starting point is located in the midpoint between the response thresholds and thus reflects no bias towards either of the response alternatives. However, if $z > {}^{a}/_{2}$, then the starting point tends towards the upper threshold, reflecting a prior bias towards the response option that is represented by that threshold. Similarly, $z < {}^{a}/_{2}$ reflects a prior bias toward the response option represented by the lower threshold.

Dynamic bias, on the other hand, is implemented by enhancing the drift rate in favor of one of the alternatives, by adding a constant v_c[8] to the drift. For example, consider a trial with drift rate v (prior to dynamic biasing). If the trial corresponds to the upper "positive" threshold, then its total drift rate (including the influence of the dynamic bias) is $v + v_c$, whereas if the trial corresponds to the lower "negative" threshold, its total drift rate (towards the negative threshold) is $v - v_c$. In the following analysis, I will assume that the upper positive threshold represents the more *a priori* likely alternative. Thus, a positive value of v_c reflects a dynamic bias towards the more *a priori* likely alternative.

Optimality of decision rules

The relationship between Wald and reward rate optimality

A *decision rule* specifies at each moment in time whether an additional sample of evidence should be taken, or else the collection of information (for the current trial) is terminated and one of the two response alternatives is selected. To recapitulate, WO (Wald optimality) is achieved by minimizing MRT (across all possible decision rules) while achieving a desired level of accuracy, or equivalently, by maximizing accuracy for a mean decision time. Denote by $AC_{Wald}(t_d)$ the maximal achievable accuracy level for a mean decision time t_d. By definition, $AC_{Wald}(t_d)$ is the accuracy achieved by a Wald-optimal decision rule (with mean decision time t_d).

RRO (reward rate optimality) is maintained by maximizing the reward rate (RR). Here, I define reward rate of a decision rule as:

$$RR = \frac{AC}{t_d + t_{res}}, \tag{1}$$

where AC denotes the expected accuracy level of the decision rule, t_d denotes the mean decision time of the decision rule and t_{res} denotes the "residual time" that includes all nondecision components that contribute to the lag between consecutive trials, e.g., the motor response production time and an inter-trial interval. It is assumed that the decision rule affects only AC and t_d but not t_{res}.[9] Denote by $AC_{RR}(t_{res})$, $t_{d,RR}(t_{res})$ respectively the accuracy and mean decision time that are achieved by the RR-optimal decision rule, for a given environment and a mean residual time t_{res}.

Appendix A shows that these two forms of optimality are equivalent in the following sense. First, if a decision rule maximizes the RR, then it is also a WO rule (Bogacz et al., 2006; Bogacz, 2009). In other words, $AC_{RR}(t_{res})$ must be the maximal level of accuracy that is possible given the mean decision time $t_{D,RR}(t_{res})$. Thus,

$$AC_{RR}(t_{res}) = AC_{Wald}\left(t_{D,RR}(t_{res})\right), \tag{2}$$

Second, given a mean value of the decision time t_d and its associated WO accuracy $AC_{Wald}(t_d)$ there exists a value of mean residual time t^*_{res} such that Eq. 2 holds. In other words, any WO combination of accuracy and mean decision time is RRO, for an appropriate value of the residual time.

The upshot of the current discussion is that the two forms of optimality are equivalent in that the same decision rules achieves both types of optimality. Therefore, by identifying the optimal strategy with respect to one of these definitions of optimality, one also can identify the optimal strategy for the other optimality type. For example, the problem of identifying the WO strategy for some t_d can be reduced to the equivalent problem of identifying the RRO strategy for the same environment and for the appropriate t^*_{res}. This problem, in turn, can be solved with the dynamic programming approach (Drugowitsch et al., 2012).

Optimality in homogeneous environments

The free-RT task When the environment is homogeneous, the SPRT (Wald and Wolfowitz, 1948; Laming, 1968) model achieves WO (and hence also RRO for an appropriate t_{res}). In this model, the decision variable corresponds to the logarithm of the likelihood ratio (LR) between the alternatives. The log-LR is updated online according to the Bayes rule as perceptual samples are incoming. Biased environments (e.g., 75 % of the correct decisions are "A" rather than "B") are

[7] This in effect assumes no integration costs. When there are such costs, integration may terminate prior to the interrogation time T (see Drugowitsch et al., 2012). See also Footnote 2.

[8] Here I make the assumption that the dynamic bias is time-constant and hence that the integration bias, $\frac{a}{2} - z + v_c t$, builds up linearly during the trial. More generally, v_c could be a function of time, but I do not consider this possibility here.

[9] Some formulations of the reward rate assume that errors are followed with negative-reward penalties and/or an increase in the inter-trial temporal interval. Here, for simplicity, I assume that no such penalties exist.

easily accommodated into the model without sacrificing optimality. This is achieved by adopting a prior bias (i.e., the "starting point"- the initial likelihood ratio from which integration of perceptual evidence begins) that reflects the external bias between the response alternatives (Edwards, 1965). Furthermore, in such (homogeneous and biased) environments, the DMM is equivalent to the SPRT (Gold and Shadlen, 2007; see Bitzer, Park, Blankenburg, & Kiebel, 2014 for the equivalence between the DDM and Bayesian models). In conclusion, the DDM with a prior bias is both the WO and RRO decision rule.

One caveat should be mentioned. In some applications of DDM (e.g. Ratcliff & Mckoon, 2008), across-trial variability parameters in the drift rate and in the starting point are accommodated. Importantly, the SPRT is equivalent to the DDM without such across-trial variability (Gold and Shadlen, 2007). When such sources of variability are introduced into the model, DDM is no longer equivalent to SPRT and hence, the DDM ceases to be optimal.

Furthermore, across-trial random variability in drift rates can result from either a subjective fluctuation in the level of psychological variables, such as attention and alertness, even when the environment is homogeneous, or from a real, objective heterogeneity of the environment. Because these two cases are indistinguishable from the viewpoint of the DDM, in the current paper I attribute all variability in drift-rate to environmental-heterogeneity. In other words, any form of drift rate variability is subsumed under the following section (Optimality in heterogeneous environments), in which heterogeneous environments are discusses. In particular, when a single objective difficulty level (homogeneous environment) is paired with subjective variability of drift rate across trials, the optimal decision rules can be identified by studying the equivalent heterogeneous environment, in which the underlying variability in drift-rate is considered to be objective.

The interrogation task Because the integration time is under the control of the experimenter in interrogation tasks (see Footnote 7), the optimal strategy (both WO and RRO) consists of updating the log-LR until the response signal is issued (rather than until a criterion level is met) and selecting the more likely alternative. Equivalently, this strategy could be implemented by updating a diffuser $x(t)$ until the response signal is issued and deciding based on the sign of $x(t)$. An external environmental bias is accommodated in the same way as in free-RT tasks: a prior for the log-LR calculation or a starting point for the diffuser.

Optimality in heterogeneous environments

The free-RT task Crucially, when the environment is heterogeneous optimality of SPRT or the DDM ceases to be the case. In such cases, optimality (RRO or WO) is achieved by integrating information according to a diffusion-like dynamics while adjusting the choice-threshold during the course of a trial (Drugowitsch et al., 2012). Equivalently, the integration process could be described as an online calculation of the log LR for the correctness of the two response alternatives and maintaining a temporally flexible threshold on the log LR. The precise form of this threshold adjustment (either in terms of log-LR or in terms of perceptual evidence) is found with the aid of a dynamic programming method. Importantly, the optimal strategy does not terminate when the total amount of integrated evidence $x(t)$ (as in DDM) or the log-LR (as in SPRT) has reached a fixed threshold (see Drugowitsch et al.). Consequently, neither SPRT nor DDM are optimal.[10]

It also is instructive to note that unlike homogeneous environments, in heterogeneous environments, the SPRT is no longer equivalent to the DDM. In this case, the SPRT is equivalent to a diffusion model with temporally increasing threshold separation. Indeed, when environments are heterogeneous the decision variable of the DDM (i.e., the total amount of accumulated evidence) no longer determines uniquely the posterior likelihood ratio for the choice alternative (see Footnote 3 for the difference between evidence and the LR). Thus, SPRT and DDM are no longer equivalent. Rather, the likelihood ratio depends also on the integration time. This fact is proved formally in Appendix B, so here I only sketch an intuition (see also Kiani & Shadlen, 2009; Drugowitsch et al., 2012).

First, for homogeneous environments, the transformation between accumulated evidence (DDM) to likelihood ratio (SPRT) depends on the difficulty level (drift rate). For example, if $x(t)=c$, then the likelihood ratio at time t is an increasing function of the drift rate (when c is held at a constant level). Equivalently, the higher the difficulty level, the higher the amount of evidence that is needed to achieve a target level of likelihood ratio at time t. Second, when the environment is heterogeneous, easy (high drift) trials tend to terminate earlier than hard (low drift) trials. Hence, as time unfolds and a trial is (still) undecided, likelihoods for high rather than low difficulties increase. This indicates that with the passage of time, a higher "conversion rate" between evidence and likelihood ratios should be adopted. In conclusion, the longer into an undecided trial, the more the threshold separation should increase (if the desired level of likelihood is to be obtained), reflecting the increasing likelihood of high difficulties. Alternatively, if a constant choice threshold is maintained then the likelihood ratio, when the threshold is reached, is a monotonically decreasing function of the threshold-reaching time.

[10] For a homogeneous environment, no threshold adjustment is necessary according to the dynamic-programming based decision rule (See Drugowitsch et al., 2012) and hence both the DDM and the SPRT are optimal.

Let me now make the "ideal observer" assumption that participants are fully aware of the prior mixture of difficulties in the heterogeneous environment. In this case, rather than simply accumulating evidence, participants may utilize the Bayes rule to update an online calculation of the likelihood ratio and decide when that ratio achieves a criterion level—the SPRT model.

Consider the "Gaussian environment" case where the prior difficulty is distributed according to a Gaussian. I assume that across trials $v \sim N(\pm v_0, \eta^2)$ where v_0 corresponds to the mean drift rate and η (dubbed drift rate variability), to the standard deviation of the Gaussian drift rate distribution. The ± corresponds to the two response alternatives. It is assumed that the observer knows the parameters v_0, η^2. The task is to identify, based on an incoming stream of evidence, the correct alternative that is, whether the current-trial drift rate was generated from the $N(v_0, \eta^2)$ or from the $N(-v_0, \eta^2)$ distribution.

As shown in Appendix B, in this case the SPRT can be implemented by the following mechanism: observers integrate evidence with a diffuser, but adopt a linearly temporally increasing threshold separation, rather than constant (i.e., time invariant) thresholds. Specifically, participants respond as soon as $x(t) = z - \frac{\alpha(s^2 + \eta^2 t)}{2v_0}$, corresponding to the lower decreasing threshold, or $x(t) = z + \frac{\alpha(s^2 + \eta^2 t)}{2v_0}$ corresponding to the upper increasing threshold. Here, $\alpha = \log\left(\frac{A}{1-A}\right)$ where A is the desired level of accuracy (here I assume that the environment is unbiased. For the general biased case, see Appendix B).

In summary, in heterogeneous environments integration up to a constant level of the likelihood ratio (as in SPRT) is equivalent to a diffuser integrating to threshold, where threshold separation is temporally increasing, rather than constant as in the DDM. To conclude this section, note that the Gaussian environment (studied above) is indistinguishable from a homogeneous environment paired with subjective Gaussian drift rate variability across trials. Thus, given subjective variability in drift rate, SPRT is no longer equivalent to DDM, even for a homogeneous environment (see Optimality in homogeneous environments).

The interrogation task For interrogation tasks, matters are simpler. The optimal strategy is to integrate the log-LR (taking into account the distribution of difficulties) and to choose the more likely alternative upon arrival of the response signal.

Appendix B shows how this log-LR is calculated for Gaussian environments. This calculation could be performed based on the diffusion dynamics $x(t)$ with $z=0$. According to Eq. (B7)

$$\widetilde{\pi} = \frac{2x(t)v_0}{(s^2 + \eta^2 t)} + \pi, \tag{3}$$

where $\pi, \widetilde{\pi}$ are the prior and posterior log-LR of both alternatives respectively and the drift rate is distributed $\sim N(\pm v_0, \eta^2)$ (the ± correspond to the "positive" and "negative" alternatives).

Interestingly, optimal behavior is based on a temporally decreasing *amplification* of the total accumulated evidence by a factor of $\frac{2v_0}{(s^2+\eta^2 t)}$. The intuition behind this finding is as follows: consider two cases where a given level of accumulated evidence say $x(t) = 1$, is collected by time $t = 10$ or alternatively by $t = 1$. The former case is more indicative of a low drift rate than the latter. The lower the drift rate the less informative the evidence is relative to the prior bias π (because the perceptual signal to noise ratio decreases). Hence, $x(t) = 1$ should be given less weight in the former case. In conclusion, the amplification of the evidence should decrease monotonically as a function of t.

The upshot of this discussion is that for *biased* heterogeneous environments, decisions are optimally based on the sign of the posterior log-LR $\widetilde{\pi}$, which is not always identical to the sign of $x(t)$. In other words, deciding based on the sign of $x(t)$ is suboptimal (once again, this is also true for homogeneous environments with subjective drift rate variability). The amplification mechanism is necessary to attain optimality. When the environment is unbiased, however, $\pi=0$ and hence $x(t)$ and $\widetilde{\pi}$ are identical in sign. Therefore, for the unbiased case, choosing based on the sign of $x(t)$ yields optimality.

Optimality restricted to the DDM

Rather than trying to identify the generally optimal decision rule (when all potential decision rules are considered), I now focus on a narrower, more modest, question. I assume that the processing architecture implements the DDM and ask, what is the optimal choice of the strategic/controllable model parameters (including prior and a dynamic bias)? Note that this is not a question of general optimality, but rather a question of *optimality restricted to the DDM*. Of focal interest is the conjecture made by Hanks et al. (2011): is the optimal dynamic bias parameter indeed positive?

The free RT task

A recent analysis by van Ravenzwaaij et al. (2012) provided a negative answer for the above question, thus supporting the conclusion that optimality in DDM is obtained with the aid of prior bias alone. Unfortunately, these analysis and conclusion require revision, as I now show. Incidentally, it is interesting to note that in their analysis van Ravenzwaaij et al. assume that the DDM is the generally optimal decision rule for heterogeneous environments. This explains the fact that these authors only focused on identifying the best selection of bias parameters within the framework of DMM. However, this attempt bypasses the question that was central in the first part of the current paper, whether DDM provides an optimal decision-mechanism in the first place. The answer is negative, as we now know. Consequently, the van Ravenzwaaij et al. study is better described as a study of optimality limited to the DDM, rather than general optimality.

van Ravenzwaaij et al. (2012) considered a given set of diffusion parameters: the external bias across alternatives, i.e., the *a priori* probability that the alternative corresponding to the positive threshold is correct (denoted β), the threshold separation a, the diffusion noise s, and a Gaussian distribution of drift rates $v \sim N(\pm v_0, \eta^2)$ (corresponding to a mixture of difficulty levels). Correctness of choices was determined according to the generative distribution of the drift v for a given trial (i.e., the "upper" threshold is the correct choice if and only if the drift on a given trial was generated according to $N(v_0, \eta^2)$ rather than $N(-v_0, \eta^2)$). Next, they determined a desired level of accuracy (e.g., 90 % or 95 %). Finally, they identified the prior and dynamic bias parameters, z and v_c that minimize MRT while maintaining accuracy at the desired level.

Whereas such an analysis could reveal the "optimal" bias parameters for the chosen threshold separation a, it overlooks the fact that a also is a free parameter that is under the control of the participant. Therefore, the search for the optimal parameters should be conducted over a three- rather than a two-dimensional parameter space (but see General discussion for a different view and for further discussion of this issue). The three dimensions consist of the prior bias z, the dynamic bias v_c, and the threshold separation a. The desired level of accuracy defines an iso-accuracy surface (in this three-dimensional parameter space) and the optimal triplet of parameters minimizes MRT on this surface.

Instead, van Ravenzwaaij et al. (2012) chose an arbitrary value of the response threshold a and then searched for a combination of biases that minimized MRT for that value of a (while maintaining accuracy at the desired level). They found that for the optimal pair of biases, the dynamic bias was zero. Hence, they concluded that optimality is obtained by the sole incorporation of a prior bias. To recapitulate, the limitation in this approach is that the decision threshold a also should be optimized as part of the search in the parameter space and not be set to an arbitrary value. The oversight in the approach of these authors is far from trivial, because it is not obvious from the outset that qualitative conclusions about the biases depend on a. But as I show below, they do.

There was an additional oversight in the analysis of van Ravenzwaaij et al. (2012), which deserves mentioning, for the benefit of future researchers. When searching for the best combination of biases, the authors considered only combinations with a positive dynamic bias (Wagenmakers, personal correspondence). However, there are also bias combinations with negative values of the dynamic bias that achieve the target level of accuracy. It is thus important to consider these combinations as well, when minimizing MRT. Interestingly, when I repeated the analysis for the same (nonoptimal) value of choice threshold a that the authors considered, I found better (i.e., lower MRT) bias combinations with negative dynamic biases! By focusing on the theoretical prediction by Hanks et al. (2011), who specifically postulated a positive dynamic bias, van Ravenzwaaij et al. overlooked the possibility that dynamic bias can be negative and hence they failed to find the optimal parameter combination (Wagenmakers, personal correspondence). Had they considered negative values of dynamic bias as well, they would have noticed that their bias combination falls short of optimality.

Thus, I reanalyzed the examples that van Ravenzwaaij et al. (2012) considered in their paper with the following changes: 1) mine was a three-dimensional (rather than two-dimensional) search, and 2) the dynamic bias was unconstrained, so the search also included negative values of dynamic bias (full details of the analysis are provided in Appendix C). The results are strikingly different than those that were reported by van Ravenzwaaij et al.

Consider the following example (studied in van Ravenzwaaij et al., 2012): the drift rate is distributed across trials $N(0.3, 0.1^2)$, the diffusion noise is $s = 0.1$, the external choice bias is $\beta = 0.8$, and the desired accuracy level is $A = 95\%$. The current results are displayed in Fig. 1 (compare with the right panels in Ravenzwaaij et al., Figure 7). The top panel depicts the tradeoff between the dynamic bias (the ordinate) and the prior bias (as a percentage of the threshold separation a; the abscissa) that generate the desired accuracy level. Note that this

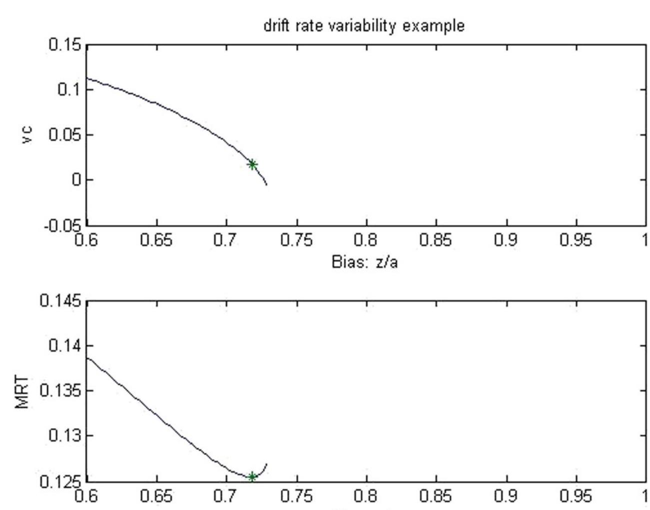

Fig. 1 The optimal biases for the Gaussian drift rate distribution example. The top panel depicts the tradeoff of prior bias (in percentage of decision threshold; abscissa) and dynamic bias (ordinate) that maintain accuracy at the desired 95 % level, when the response threshold is at its optimal level $a = 0.111$. The bottom panel depicts MRT as a function of prior bias (accompanied by the corresponding dynamic bias). The asterisk corresponds to the optimal performance

tradeoff is depicted for threshold separation a = 0.111. This value was not selected arbitrarily but rather it was obtained as the a coordinate of the optimal parameter triplet (this value of a is different from the value considered by van Ravenzwaaij et al.).

The bottom panel of Fig. 1 depicts the MRT as a function of the prior bias (when it is accompanied with the corresponding "top panel" dynamic bias), for the same value a. The asterisk, which is the point of minimum MRT, corresponds to optimal performance. As can be read from the top panel, the optimal bias combination features a positive dynamic bias parameter $v_c \cong 0.02$ in addition to a positive prior bias z. Crucially, in violation of the conclusion of van Ravenzwaaij et al., optimality in DDM mandates accommodating a positive dynamic bias component in the integration bias (in addition to prior bias). Interestingly, had van Ravenzwaaij et al. considered other values of the threshold separation a they would have found (e.g., for a = 0.11 as well as for other values), that the "two-dimensional optimal dynamic bias" (i.e., for this specific value of a, when only the bias-parameters are optimized) is positive.

To probe the issue further, I simulated other scenarios in which the drift rate distribution was discrete rather than Gaussian (i.e., across trials the drift rate was drawn from a finite set of values). In one such example, I mixed two drift rates $v \in \{0.02, 0.05\}$, assuming that each difficulty level appears in 50 % of the trials and that the same external bias β = 0.65 applies to both difficulty levels. Setting the desired accuracy level at A = 80%, I searched for the optimal (a,z, v_c) triplet. Here, I obtained $v_c \cong 0.024$—a bias that is comparable in magnitude to the nonbiased drifts! The upshot of this example is that in such cases too (discrete rather than Gaussian mixtures of difficulties), optimal combinations of biases (and threshold separation) feature a positive dynamic bias. Thus, this discrete mixture case also provides support for the conjecture of Hanks et al. (2011) and contradicts the conclusion of van Ravenzwaaij et al. (2012).

Can a negative DDM dynamic bias be optimal? The analysis thus far suggests one puzzling question: is it possible that for some biased heterogeneous environments the optimal dynamic bias in DDM is negative? This question is interesting, because such environments would contradict the Hanks et al. (2011) conjecture that postulated a positive dynamic bias. Therefore, a treasure of novel knowledge about the mechanisms of integration biases may be exposed if such environments are discovered.

Importantly, all the examples of heterogeneous environments that were explored in the current paper (as well as in several other examples that I simulated and that are not reported here) yielded a *positive* dynamic bias. Nevertheless, in some of these examples when I manually fixed the threshold separation a to a value that was much higher than optimal, and searched for the "best" pair of bias parameter for that arbitrary a—the "best" dynamic bias was negative. To reiterate, this bias is nonoptimal, because a should also be part of the optimization. Nevertheless, this result highlights the question of generality with respect to the positivity of the dynamic bias in DDM. I leave this question open for future study, which would either prove the generality of this result or identify puzzling contradicting examples.

The interrogation task

In studying the interrogation task, I stress once more that the globally optimal strategy for Gaussian heterogeneous environments, which is based on log-LR calculations and described in the first part of the paper, involves a multiplicative temporally decreasing amplification of the accumulated evidence rather than using an additive dynamic bias. In the current section, I explore the optimal strategy that is feasible with the aid of a prior and an additive dynamic bias alone. In this case, decision is based on the sign of the total accumulated evidence (including the prior and dynamic bias components).

In analyzing the Gaussian environment ($v \sim N(\pm v_0, \eta^2)$), van Ravenzwaaij et al. (2012; p. 6) showed that the optimal pair of prior (z) and dynamic (v_c) biases must satisfy the equation:

$$z + v_c T = \frac{(s^2 + \eta^2 T)\log\left(\frac{\beta}{1-\beta}\right)}{2v_0}, \tag{4}$$

where T is the interrogation time. These authors noted that for a given T, z, and v_c tradeoff in that various combinations of these parameters satisfy Eq. 4. One such combination is obtained by setting the dynamic bias to zero $v_c = 0$, and the prior bias, z to the right hand side of Eq. 4. Thus, optimal behavior in the DDM can be achieved with the aid of prior bias alone.

Critically, the validity of this conclusion is limited to the case where the interrogation time T is fixed to a known constant value across trials. What are the optimal prior and dynamic biases when the interrogation time T varies randomly across trials? The key insight in finding the optimal decision in this case is noting that, by an appropriate selection of a pair of biases, Eq. 4 can be solved *simultaneously* for all values of T. Indeed, both sides of Eq. 4 are linear in T. These two lines coincide and equality holds for all values of T if and only if:

$$z = \frac{s^2 \log\left(\frac{\beta}{1-\beta}\right)}{2v_0}, \tag{5}$$

$$v_c = \frac{\eta^2 \log\left(\frac{\beta}{1-\beta}\right)}{2v_0}, \tag{6}$$

Using this pair of biases (Eqs. 5–6), guarantees that for any T that realizes on a given trial, the participant behaves optimally, as if T was known in advance! On the other hand, if an alternative pair of biases is selected, Eq. 4 will be violated for at least one value of T (as long as more than a single value of T is intermixed in the block).[11] For such Ts, the violation of Eq. (4) manifests in a lower accuracy level relative to the accuracy level that is achieved with the Eqs. (5–6) pair of biases, which satisfies Eq. 4. Therefore, any alternative pair of biases is inferior to the optimal pair defined by Eqs. 5–6. Notably, to achieve optimality, the observer does not even have to know how T distributes across trials, because the optimal pair of biases is invariant with respect to this distributions.

Finally, according to Eq. 6, $v_c > 0$ (assuming that $\beta > 0.5$). The upshot is that when interrogation time varies randomly across trials, a positive dynamic bias is mandatory for achieving optimality, in accordance with the conjecture of Hanks et al. (2011).

General discussion

When biased environments are homogeneous with respect to difficulty level, optimal performance is achieved by the sequential probability ratio test (SPRT) with a prior bias (Edwards, 1965) or equivalently by the drift diffusion model (Gold and Shadlen, 2007). Unfortunately, shifting to heterogeneous environments complicates matters immensely. The current analysis shows that: 1) reward rate-optimality is equivalent to Wald optimality; 2) in heterogeneous environments, the SPRT is no longer equivalent to the DDM model, because the likelihood ratio is a function of both the total accrued evidence (the decision variable of the DDM) and the integration time. Rather SPRT is equivalent to a diffusion model with temporally increasing threshold separation; and 3) in heterogeneous environments, both the SPRT and DDM are suboptimal, even for unbiased environments. Rather, optimality is achieved with the aid of a diffusion-like integration of information (or by updating log-LR calculations), while adjusting the choice threshold, within the course of a trial (Drugowitsch et al., 2012).

Next, I restricted focus to the issue of identifying optimal performance within the framework of the highly popular and successful DDM. Because this model provides an accurate account for decision making in a wide range of choice-tasks, it is instrumental to study the question of optimality under its auspices. Specifically, I attempted to identify the optimal ensemble of boundary separation, the prior bias and the dynamic bias parameter-triplet, given the features of the environment (the external bias and the distribution of drift rates). I found that optimality in a free RT task requires a non-zero dynamic bias parameter. That is, rather than being time-invariant, the integration bias builds up with the passage of time. This conclusion is consistent with the conjecture, that in heterogeneous biased environments, both a prior and a dynamic bias are necessary if optimality is to be achieved (Hanks et al., 2011). These conclusions also contradict prior conclusions by van Ravenzwaaij et al. (2012), according to which optimality is achieved with the aid of prior bias solely. Additionally, while a dynamic bias is not necessary to achieve optimality in interrogation tasks that feature a single known interrogation time, it is mandatory, when interrogation time varies randomly across trials. The conjecture by Hanks et al. thus seems viable, at least as long as DDM is the decision mechanism.

Psychological plausibility of optimal models

The discussion thus far was purely theoretical in that it studies optimal performance disregarding the question whether such (optimal) behavior is achievable. Next, I discuss briefly the psychological plausibility of the optimal algorithms/strategies that were presented in the current paper. Can the mathematically optimal models transcend the mathematical realm and take the form of a psychological reality?

One unreasonable assumption casts doubt on the plausibility of the optimal strategies that have been discussed thus far. The mathematical implementations of the optimal strategies take advantage of perfect and full knowledge of the distribution of drift rates across trials. In other words, these implementations assume that observers are omniscient with respect to the statistical properties of the environment.

While the assumption that human observers may have fairly accurate (albeit not perfect) statistical representations in highly trained and familiar environments may be reasonable, this assumption is certainly invalid for many of the novel and unfamiliar tasks that participants encounter in the cognitive laboratory. Thus, it is more realistic to expect participants to approximate optimality, only following a training period during which participants learn to represent the environment. Conceivably, as people experience with a task, they form a gradually improving representations of the statistical features of the environment, which in turn, allow them to execute increasingly successful choice strategies. Consistent with this idea, several algorithms, which learn to represent probability distributions functions have been presented in the literature (see Turner, Van Zandt and Brown, 2011, for a dynamic stimulus driven model for signal detection tasks, and see Rao, 2004, for a recurrent network architecture).

[11] The two linear functions of T, in both sides of Eq. 4 either coincide (for biases that are selected according to Eqs. 5–6) or otherwise intersect for at most a unique value of T.

Do the optimal models become psychologically plausible once observers have faithfully represented (following extensive learning) the statistical properties of the environment? Even if we assume that an observer is omniscient with respect to the environment, we soon bump into the computation hurdle. To illustrate, my own calculations of the optimal DDM parameter-triplet and let alone, the Drugowitsch et al. (2012) calculations of the optimal threshold adjustment policies, all relied on highly complex computational algorithms (i.e., dynamic programing and a simplex search in a parameter space). Can we realistically assume that human observers have such complicated algorithms at their cognitive disposal?

Consider first the more modest issue of optimality restricted to the DDM model. One possibility that bypasses altogether the need to form complex statistical representations of the environment or to execute extremely complex computations may be reinforcement learning. Participants may use an adaptive adjustment procedure: different parameter triplets are sampled, the effects on performance are then observed across numerous trials (e.g., the reward rate or the ensuing speed-accuracy combination) and further parameter adjustments are implemented (e.g., according to a gradient-based reinforcement learning).

Myung & Busemeyer (1989) and Busemeyer & Myung (1992) showed that by applying such an algorithm participants achieve close convergence (albeit slowly) of their choice threshold to its optimal value.[12] Admittedly, DDM-optimality in heterogeneous and biased environments is computationally more complex as observers should optimize on a triplet of parameters, rather than on a single parameter. Nonetheless, reinforcement learning algorithms warrant optimism with respect to the plausibility of converging on a triplet that comes close to optimal.

Shifting focus to the threshold adjustments algorithm of Drugowitsch et al. (2012), it is more difficult to see how adaptive adjustment procedures can approximate optimality. The reason is that the time-course of the threshold should be adjusted continuously during a trial. Suppose that people approximate such a continuous adjustment by adjusting their threshold every temporal interval Δt. On the one hand, Δt should be small if a close approximation to the optimal continuous adjustment procedure is to be achieved. On the other hand, a minute Δt manifests in a highly dimensional parameter space.[13] Because the complexity of parameter-space search algorithm inflates as the parameter-space dimensionality increases, effective searches in the parameter space become enormously time-costly. Future studies should explore whether human observers can approximate optimality in their threshold adjustments policies (assuming that they can apply threshold adjustment policies in the first place) and if they can, which algorithms guide such adjustments.

Can people optimize on their choice threshold?

The model of Drugowitsch et al. (2012) assumes that observers optimize on their choice threshold by adjusting it during the time course of a trial. In studying optimality restricted to DDM, I made the much weaker assumption that observers optimize on their choice threshold by setting it to a constant level (both within and across trials) that is tailored to the statistical properties of the environment. An alternative view posits that, perhaps even this assumption is too liberal and that it may be more psychologically plausible to assume that participants cannot or are unwilling to optimize on their choice threshold (Don van Ravenzwaaij, personal correspondence). For example, the choice threshold may provide certainty about the identity of the stimulus and observers may set the choice threshold to attain a desired target level of certainty. Thereafter, observers try to maintain optimality in heterogeneous and biased environments by adjusting only the pair of prior and dynamic biases.

Interpreted in this light, the analysis of van Ravenzwaaij et al. (2012) identifies the optimal strategy, under the assumptions that DDM is the processing model and that the *a* parameter is excluded from the optimization (note however, that the possibility of a negative dynamic bias was overlooked). If optimality is thus defined, my conclusions with respect to the dynamic bias alter: The optimal dynamic bias can obtain positive but also zero and even negative values for different settings of the threshold in a given environment (see section Optimality restricted to DDM). The last two cases conflict with the conjecture of Hanks et al. (2011), which envisioned a positive dynamic bias.

I argue, however, that from a psychological perspective it is more instrumental to relieve this constraint. The perceptual-choice literature consists of ample evidence that the choice threshold is under the cognitive control of people, and that participants can and do adjust their choice thresholds across different blocks of trials as a function of the experimental instructions (stressing speed choice or speed accuracy). Furthermore, as discussed earlier, there is evidence that participants can optimize on their choice threshold via reinforcement learning (Myung & Busemeyer, 1989; Busemeyer & Myung, 1992; see also the section *Threshold Setting Algorithms* in Bogacz et al., 2006). Finally and perhaps most relevantly, observers adjust their choice threshold as a function of the distribution of stimulus-difficulty in the environment (e.g., the *blocking effect*; Mozer, Kinoshita, & Davis, 2004). Therefore, I find no compelling reason to exclude *a priori* the choice threshold from the optimization parameter set, while keeping the pair of biases in.

[12] These studies used a different criterion for optimality namely, the Bayes Risk (BR) which minimizes a weighted sum of the mean RT and Error rate (Wald & Wolfowitz, 1948).

[13] For example, if an observer adjusts his or her threshold every 100 ms during a two second interval then 20 parameters are required to describe the adjustment procedure.

In summary, my perspective is that intransigent post-training deviations between the optimal and the actual choice thresholds represent failures of optimality. If empirical data reveals that such failures occur, then the question of *why* they occur emerges. A variety of potential answers should then be considered. For example, the choice threshold could optimize a psychological goal that is different from the goal, which is subsumed under the focal optimality criterion (e.g., it satisfies a desired certainty about the identity of the stimulus, a goal that is absent from the definitions of WO or RRO). Another possibility is that participants use parameter adjustment (reinforcement learning) algorithms, but they fix their choice threshold at a constant level to decrease the complexity (the dimensionality of the search space) of the adjustment process. These are only two of the numerous potential causes for optimality failure (is it possible that the observer is not relying on a diffusive integration of information?).

Conclusions

The current discussion highlights some of the reasons for the prominence of the concept of optimality in guiding the study of decision making. Optimality serves as an important standard, benchmark, and yardstick in evaluating human choices. When people are found to perform optimality, probing their behavior advances our understanding with respect to *how* their (optimal) behavior is produced, for example, which processing algorithms are executed and/or which adaptation mechanisms are used. Hopefully, such principles can be leveraged and applied in additional spheres of behavior. However, because organisms are generally "expected" to adapt efficiently to their environment, deviation from optimality may sometimes be even more striking than compliance with optimality (Tversky & Kahneman, 1974, for an example from a different domain: The "Heuristics and Biases" approach). When agents violate optimality, we must account not only for *how* behavior is produced but also for *why* such behavior is produced (Norris, 2006, 2009). In other words, we face the question pertaining to the causes underlying non-optimality. This inquiry, in turn, paves the way for a plethora of follow-up research directions: did we misidentify the agent's goal when in fact, she is optimizing on a different goals? Or perhaps we identified correctly the agent's goals, but she is limited in her resources or in her ability to represent faithfully the environment? And so on and so forth. The pursuit of answers for such questions is a powerful driving force, contributing to the advancement and to a refinement of our understanding of behavior.

Author Notes The author thanks Marius Usher (MU), Eric-Jan Wagenmakers (EJW), and Don van Ravenzwaaij (DVR) for helpful discussions, and also EJW, DVR, and Andrew Heathcote for providing excellent suggestions during the revision process, and finally, Konstantinos Tsetsos for comments about an earlier version of this manuscript.

Appendix A: Relationship between Wald and RR optimality

In this appendix, I show that WO and RRO are equivalent. This means that a decision rule which achieves one form of optimality achieves also the other, as specified below.

An RRO decision rule is also WO

In this section, I show that if a decision rule maximizes the reward-rate then it is also a Wald optimal strategy (Bogacz et al., 2006; Bogacz, 2009). Denote by $\widetilde{D}$ a decision rule that achieves a maximal reward rate for a given environment and for a given value of mean residual time t_{res}. The accuracy and mean decision times for $\widetilde{D}$ are $AC_{RR}(t_{res})$ and $t_{d,RR}(t_{res})$ respectively. I argue that $\widetilde{D}$ is also Wald optimal in that $AC_{RR}(t_{res})$ must be the maximal possible accuracy (among all decision rules) with mean decision time $t_{d,RR}(t_{res})$. That is:

$$AC_{RR}(t_{res}) = AC_{Wald}\left(t_{d,RR}(t_{res})\right), \tag{A1}$$

To see this, note that the reward rate that a WO decision rule, with mean decision time $t_{d,RR}(t_{res})$ achieves is given by $\frac{AC_{Wald}\left(t_{d,RR}(t_{res})\right)}{t_{d,RR}(t_{res})+t_{res}}$. Since the Wald rule provide the maximal accuracy for a given mean decision time,

$$AC_{Wald}\left(t_{d,RR}(t_{res})\right) \geq AC_{RR}(t_{res}), \tag{A2}$$

Thus the reward rate of the Wald optimal rule is:

$$\frac{AC_{Wald}\left(t_{d,RR}(t_{res})\right)}{t_{d,RR}(t_{res})+t_{res}} \geq \frac{AC_{RR}(t_{res})}{t_{d,RR}(t_{res})+t_{res}} = RR\left(\widetilde{D}\right), \tag{A3}$$

In words, the RR of the Wald optimal rule (with mean decision time $t_{d,RR}(t_{res})$) is at least as large as the reward rate obtained by $\widetilde{D}$. However, by definition $\widetilde{D}$ is the RR-optimal rule and therefore an equality must hold in Eq. A3 and thus in Eq. A2 as well. Hence Eq. A1 is satisfied.

WO decision rule is also RRO

In the current section I show that, given a target mean value of the decision time t_0 there exists a positive mean residual time

t^*_{res} for which the WO rule (with mean decision time t_0 and its associated Wald-optimal accuracy $AC_{Wald}(t_0)$) maximizes the reward rate. To simplify notation, henceforth I denote the mean decision time by t (instead of t_d) and the maximal accuracy by $A(t)$ (instead of $AC_{Wald}(t_d)$

The function $A(t)$ has several important properties. First, by taking 0 decision time an observer can achieve a maximal accuracy of $\max\{p_1, 1-p_1\}$ where p_1 is the a-priori probability that option '1' (rather than '2') is correct. Without loss of generality we can assume that $p_1 \geq 0.5$ and hence $A(0) = p_1$.

Second, $A(t)$ is a monotonically increasing function of t. Indeed, if $t_1 > t_2$ then one potential decision rule with mean decision time t_1 is to adopt the WO decision rule for a mean decision time t_2 and then 'sit and wait' for a duration of $t_1 - t_2$ before issuing a decision. This will yield accuracy of $A(t_2)$. Of course, waiting without integrating information is suboptimal because observers can collect further information which will facilitate accuracy. Therefore $A(t_1) > A(t_2)$.

Third, $A(t)$ is a concave function of t. This means that for all t_1, t_2 and $\lambda \in [0,1]$:

$$A(\lambda t_1 + (1-\lambda)t_2) \geq \lambda A(t_1) + (1-\lambda)A(t_2), \tag{A4}$$

Indeed, Consider the following 'mixture' decision rule: With probability $\lambda \in [0,1]$ the observer follows the Wald optimal decision rule for mean decision time t_1 and otherwise (i.e. with probability $(1-\lambda)$), the observer follows the Wald optimal decision rule for mean decision time t_2. This mixture rule provides accuracy that is equal to the right hand side of Eq. A4 and its mean decision time is $\lambda t_1 + (1-\lambda)t_2$. By definition, A Wald-optimal decision rule for mean decision time $\lambda t_1 + (1-\lambda)t_2$ will provide at least the same accuracy, and so Eq. A4 is satisfied. If the Wald optimal decision rule will be more efficient than the mixture rule then strict concavity is obtained (i.e. in A4. we will have strict inequality).

Assuming $A(t)$ is a differentiable function, the monotonicity and concavity properties translate too

$$A'(t) > 0, A''(t) < 0 \tag{A5}$$

Consider next the reward rate, $RR = \frac{AC}{t+t_{res}}$. We already know, from the previous subsection, that any rule that maximizes the RR is a Wald optimal strategy. Therefore, the optimal reward rate is achieved by maximizing with respect to t the reward function:

$$R(t) = \frac{A(t)}{t + t_{res}}, \tag{A6}$$

Taking derivatives with respect to t we find that

$$R'(t) = \frac{A'(t)(t + t_{res}) - A(t)}{(t + t_{res})^2} \tag{A7}$$

And the condition for stationary points is thus:

$$A'(t)(t + t_{res}) - A(t) = 0, \tag{A8}$$

Consider a target mean decision time $t_0 > 0$. I next show that there exists some positive t^*_{res} for which t_0 is a stationary point. Indeed, defining

$$t^*_{res} = \frac{A(t_0)}{A'(t_0)} - t_0, \tag{A9}$$

We note that t_0 solves Eq. A8. So it remains to be seen that t^*_{res} is indeed positive:

$$t^*_{res} = \frac{A(t_0) - t_0 A'(t_0)}{A'(t_0)} = \frac{\left(A(0) + \int_0^{t_0} A'(\tau)d\tau\right) - t_0 A'(t_0)}{A'(t_0)}, \tag{A10}$$

Noting that $A'(\tau)$ is a decreasing function of τ ($A''(\tau) < 0$ according to Eq. 5), we obtained that $\int_0^{t_0} A'(\tau)d\tau > t_0 A'(t_0)$. Thus, continuing Eq. A10,

$$t^*_{res} = \frac{A(0) + \int_0^{t_0} A'(\tau)d\tau - t_0 A'(t_0)}{A'(t_0)} > \frac{p_1 + t_0 A'(t_0) - t_0 A'(t_0)}{A'(t_0)} = \frac{p_1}{A'(t_0)} > 0, \tag{A11}$$

Next, I show that the stationary point t_0 is a maximum point. Indeed, taking another derivative from Eq. A7 we obtain that:

$$R''(t) = \frac{A''(t)(t + t_{res})^3 - 2(t + t_{res})[A'(t)(t + t_{res}) - A(t)]}{(t + t_{res})^4}, \tag{A12}$$

Evaluating Eq. A12 at the stationary point t_0 simplifies to $R''(t_0) = \frac{A''(t_0)}{t_0 + t_{res}} < 0$ which shows that t_0 is a local maximum point of the reward rate.

Next, I show that t_0 is in fact a global maximum. If we assume it is not, then Eq. A8 has another root (stationary point) at the global maximum. Thus, Eq. A8 has at least two different roots. But this means that the derivative of Eq. A8 must also have a root. Thus there exists a positive t such that $A''(t)(t + t^*_{res}) = 0$, which is impossible because $A''(t) < 0$ and $t + t^*_{res} > 0$. Therefore, t_0 must be a global maximum.

To conclude, given a target mean decision time t_0, I found a mean positive value of the residual time, t^*_{res} (Eq. A11) for which the WO decision rule (with mean decision time t_0 and accuracy $A(t_0)$) is RR-optimal.

Appendix B: SPRT in Gaussian environments

In the current Appendix, I extend the SPRT model to Gaussian heterogeneous environments. I assume that the on each trial, the difficulty level is drawn from a Gaussian distribution. The unique source of uncertainty concerns which of the two response alternatives is correct. On each temporal interval dt a new independent perceptual sample is generated and is distributed $\sim N(vdt, s^2dt)$ where s^2 is the variance rate, and v is the drift rate for the current trial. The participant needs to decide in favor of one of two hypotheses:

> H_0: The current drift rate v was generated from a $N(v_0,\eta^2)$ distribution or:
> H_1: The current drift rate v was generated from a $N(-v_0,\eta^2)$ distribution.

Importantly, the *positive* parameters v_0 and η, which corresponds to the mean and to the standard deviation of difficulty distribution respectively are known.

Denote by $\widetilde{x}(t)$ and $x(t)$ respectively the entire stream of accumulated perceptual evidence and the total accumulated evidence obtained by time t (thus $x(t)$ is simply the state of $\widetilde{x}(t)$ at time t). According to Bayes' rule the posterior odds is the product of the prior odds and the Bayes factor (BF):

$$\frac{P\left(H_0\middle|\widetilde{x}(t)\right)}{P\left(H_1\middle|\widetilde{x}(t)\right)} = \frac{P\left(\widetilde{x}(t)\middle|H_0\right)}{P\left(\widetilde{x}(t)\middle|H_1\right)}\frac{P(H_0)}{P(H_1)} \tag{B1}$$

Let us next focus on the numerator term $P(\widetilde{x}(t)|H_0)$. It can be shown (see Drugowitsch et al., 2012, Eq. 10) that conditional on a drift rate v:

$$P\left(\widetilde{x}(t)\middle|v\right) = D\left(\widetilde{x}(t)\right)e^{\frac{2x(t)v - tv^2}{2s^2}} \tag{B2}$$

Where $D(\widetilde{x}(t))$ depends on the specific stream $\widetilde{x}(t)$ but not on the drift rate v.

Reading the following derivation, throughout the section proportionality ($\propto$) denotes equality up to a multiplicative term that is invariant with respect to v_0 (and its sign) but may depend on the specific stream $\widetilde{x}(t)$. Note that, $P(\widetilde{x}(t)|H_0)$ is obtained by integrating $P(\widetilde{x}(t)|v)$ over the drift distribution. Thus:

$$\begin{aligned}
P(\widetilde{x}(t)|H_0) &\propto \int_{-\infty}^{\infty} P(\widetilde{x}(t)|v)e^{-\frac{(v-v_0)^2}{2\eta^2}}dv = D(\widetilde{x}(t))\int_{-\infty}^{\infty} e^{\frac{2x(t)v-tv^2}{2s^2}}e^{-\frac{(v-v_0)^2}{2\eta^2}}dv \\
&\propto \int_{-\infty}^{\infty} e^{\frac{\eta^2 v(2x(t)-vt)\, -\, s^2(v-v_0)^2}{2s^2\eta^2}}dv \\
&= e^{-\frac{v_0^2}{2\eta^2}}\int_{-\infty}^{\infty} e^{-\frac{(s^2+\eta^2 t)v^2-2(\eta^2 x(t)+s^2 v_0)v}{2s^2\eta^2}}dv \\
&= e^{-\frac{v_0^2}{2\eta^2}}\int_{-\infty}^{\infty} e^{-\frac{v^2-2\left(\frac{\eta^2 x(t)+s^2 v_0}{s^2+\eta^2 t}\right)v}{\frac{2s^2\eta^2}{s^2+\eta^2 t}}}dv \\
&= e^{-\frac{v_0^2}{2\eta^2}}\int_{-\infty}^{\infty} e^{-\frac{\left(v-\frac{\eta^2 x(t)+s^2 v_0}{s^2+\eta^2 t}\right)^2}{\frac{2s^2\eta^2}{s^2+\eta^2 t}}}e^{\frac{(\eta^2 x(t)+s^2 v_0)^2}{2s^2\eta^2(s^2+\eta^2 t)}}dv \\
&\propto e^{-\frac{v_0^2}{2\eta^2}+\frac{2s^2\eta^2 x(t)v_0+s^4 v_0^2}{2s^2\eta^2(s^2+\eta^2 t)}}\int_{-\infty}^{\infty} e^{-\frac{\left(v-\frac{\eta^2 x(t)+s^2 v_0}{s^2+\eta^2 t}\right)^2}{\frac{2s^2\eta^2}{s^2+\eta^2 t}}}dv
\end{aligned} \tag{B3}$$

Examining the integrand in the final term we note that it is proportional to the probability density function of a Normal distribution with mean $= \frac{\eta^2 x(t)+s^2 v_0}{(s^2+\eta^2 t)}$ and variance $= \frac{s^2\eta^2}{s^2+\eta^2 t}$. Therefore this integral is independent of v_0 so we obtain:

$$P\left(\widetilde{x}(t)\middle|H_0\right) \propto e^{-\frac{v_0^2}{2\eta^2}+\frac{2\eta^2 x(t)v_0+s^2 v_0^2}{2\eta^2(s^2+\eta^2 t)}} \tag{B4}$$

We can now derive the term $P(\widetilde{x}(t)|H_1)$ by replacing in Eq. B4 v_0 by $-v_0$ to obtain:

$$P\left(\widetilde{x}(t)\middle|H_1\right) \propto e^{-\frac{v_0^2}{2\eta^2}+\frac{-2\eta^2 x(t)v_0+s^2 v_0^2}{2\eta^2(s^2+\eta^2 t)}} \tag{B5}$$

Equations A4–A5 share the same proportion factor, hence returning to the BF, it follows that

$$\frac{P\left(\widetilde{x}(t)\middle|H_0\right)}{P\left(\widetilde{x}(t)\middle|H_1\right)} = e^{\frac{2x(t)v_0}{(s^2+\eta^2 t)}} \tag{B6}$$

Finally, taking logarithms of Eq. B1 and using Eq. B6, we obtain:

$$\widetilde{\pi} = \frac{2x(t)v_0}{(s^2+\eta^2 t)} + \pi \tag{B7}$$

where π and $\widetilde{\pi}$ are the log-prior and log-posterior odds respectively.

In SPRT integration of perceptual evidence occurs until the posterior reaches a target level $\pm\alpha, \alpha \equiv \ln\left(\frac{A}{1-A}\right)$, where A is a target level of accuracy. From Eq. B7 it follows that integration occurs until

$$x(t)\varepsilon\left\{-\frac{s^2(\alpha+\pi)}{2v_0}-\frac{\eta^2(\alpha+\pi)}{2v_0}t, \frac{s^2(\alpha-\pi)}{2v_0}+\frac{\eta^2(\alpha-\pi)}{2v_0}t\right\} \quad \text{(B8)}$$

This means that a diffuser (with starting point $x(0) = z$) will terminate all trials with the same posteriors level of $\pm\alpha$ if the time-variant response thresholds are set at distances $-\frac{s^2(\alpha+\pi)}{2v_0}-\frac{\eta^2(\alpha+\pi)}{2v_0}t$ (the lower threshold) and $\frac{s^2(\alpha-\pi)}{2v_0}+\frac{\eta^2(\alpha-\pi)}{2v_0}t$ (the upper threshold) from the starting point. Note that the lower and upper response thresholds respectively are linearly decreasing and increasing functions of time and that the boundary separation increases with rate $\frac{\eta^2\alpha}{v_0}$. Additionally, in the particular case that the environment is unbiased (i.e. $\pi = 0$) both thresholds change with equal absolute rates $\frac{\eta^2\alpha}{2v_0}$ but in opposite direction.

Another implication of Eq. B7 is that in the DDM, where integration stops when either of the (constant) thresholds is reached (located in distances $-z$ or $a-z$ from the starting point) then the log odds are: $\widetilde{\pi} = \pi - \frac{2zv_0}{(s^2+\eta^2 t)}$, for the lower boundary and $\widetilde{\pi} = \frac{2(a-z)v_0}{(s^2+\eta^2 t)} + \pi$, for the other threshold. Recall, that the logs odds are formulated in terms of the 'upper' (H_0) choice-alternative relative to the 'lower' (H_1) choice-alternative. If instead, the log odds are formulated with respect to the chosen relative to the non-chosen alternative, the log-odds for the lower threshold is obtained by flipping the sign: $\widetilde{\pi} = \frac{2zv_0}{(s^2+\eta^2 t)} - \pi$. Note that the log odds for both alternatives decrease monotonically as a function of t, tending towards the prior odds ($\pm\pi$) as $t \to \infty$.

Appendix C: Simulation methods

In this appendix I describe the method I used for finding the optimal triplet (a,z,v_c) for the DDM in biased heterogeneous environments. For a single difficulty level v, the accuracy and the MRT are given by (c.f. Eq. 8–12 in van Ravenzwaaij et al. 2012):

$$Acc(v|a,z,v_c,\beta,s) = \beta\frac{\left(e^{\frac{2a(v+v_c)}{s^2}}-e^{\frac{2(a-z)(v+v_c)}{s^2}}\right)}{e^{\frac{2a(v+v_c)}{s^2}}-1} + (1-\beta)\frac{\left(e^{\frac{2a(v-v_c)}{s^2}}-e^{\frac{2z(v-v_c)}{s^2}}\right)}{e^{\frac{2a(v-v_c)}{s^2}}-1} \quad \text{(C1)}$$

$$MRT(v|a,z,v_c,\beta,s) = \beta\left(-\frac{z}{v+v_c}+\frac{a\left(e^{-\frac{2z(v+v_c)}{s^2}}-1\right)}{(v+v_c)\left(e^{-\frac{2a(v+v_c)}{s^2}}-1\right)}\right) + (1-\beta)\left(-\frac{a-z}{v-v_c}+\frac{a\left(e^{-\frac{2(a-z)(v-v_c)}{s^2}}-1\right)}{(v-v_c)\left(e^{-\frac{2a(v-v_c)}{s^2}}-1\right)}\right) \quad \text{(C2)}$$

When the environment is heterogeneous so that drift is distributed $\sim f(v)$ I found the accuracy and MRT by integrating the corresponding terms over the distribution f. The two cases that are explored in the paper are a Gaussian and a discrete f with two equi-probable drift rates. For the Gaussian case the integration was performed by numerical integration and the discrete integration was handled by arithmetic averaging.

The optimization problem can now be formulated as

$$(a,z,v_c) = argmin\int MRT(v|a,z,v_c,\beta,s)df(v) \quad s.t \int Acc(v|a,z,v_c,\beta,s)df(v) \geq A \quad \text{(C3)}$$

where A is the desired accuracy level. Note that for the optimal triplet the constraint is always satisfied with equality, otherwise a sufficiently slight reduction to the threshold separation a would diminish MRT while maintaining accuracy above the desired level, contradicting the optimality of the triplet. Defining:

$$F(a,z,v_c,\beta,s) = \begin{cases} \int \mathrm{MRT}(v|a,z,v_c,\beta,s)\,\mathrm{d}f(v) & , \int Acc(v|a,z,v_c,\beta,s)df(v) \geq A \\ \infty & , \int Acc(v|a,z,v_c,\beta,s)df(v) < A \end{cases} \quad \text{(C4)}$$

the optimal triplet is defined by $(a,z,v_c) = argminF(a,z,v_c,\beta,s)$.

I took considerable measures to avoid local minima in the search for the triplet that minimizes F. This search was conducted with a combination of genetic algorithms and the iterative Nelder-Mead (Nelder & Mead, 1965) Simplex method, (implanted by the routines "ga," "fminsearch" available in Mathwork's MATLAB). I repeated the following steps 10,000 times. First, I minimized the objective function by running the genetic algorithm. The output triplet was then fed as the starting point for the simplex algorithms. The simplex algorithm in turn was iterated several times; each iteration started with the parameters obtained from the termination of the previous iteration.[14] This was repeated until the objective function improved by less than 1e-5 on two consecutive runs. The triplet that minimized the objective function more than 10,000 iterations of a genetic algorithm followed by a sequence of simplex iterations was considered to be the optimal triplet.

References

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics, 73*(2), 640–657.

Bitzer, S., Park, H., Blankenburg, F., & Kiebel, S. J. (2014). Perceptual decision making: Drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, 8.

Bogacz, R. (2009) Optimal decision making theories. In J. C. Dreher & L. Tremblay (Eds.), *Handbook of reward and decision making*. Elsevier.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review, 113*(4), 700–765.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory and human performance. *Journal of Experimental Psychology: General, 121,* 177–194.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *The Journal of Neuroscience, 29*(37), 11560–11571.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, *6*.

Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics, 68*(2), 194–207.

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review, 16*(6), 1129–1135.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience, 32*(11), 3612–3628.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology, 2*(2), 312–329.

Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825–837). Cambridge, MA: MIT Press.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions and reward. *Neuron, 36,* 299–308.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30,* 535–574.

[14] Typically, when the Simplex algorithm converges, the search simplex has shrunk to a small diameter. By starting a novel Simplex iteration one increases the diameter of the search simplex. Thus, the next iteration can converge to a different point.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience, 31*(17), 6339–6352.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*(5928), 759–764.

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

Mozer, M. C., Kinoshita, S., & Davis, C. (2004). Control of response initiation: Mechanisms of adaptation to recent experience. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 981–986). Hillsdale, NJ: Erlbaum.

Mulder, M. J., Wagenmakers, E. J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience, 32*(7), 2335–2343.

Myung, I. J., & Busemeyer, J. R. (1989). Criterion learning in a deferred decision making task. *American Journal of Psychology, 102,* 1–16.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*(4), 308–313.

Norris, D. (2006). The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327–357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review, 116*(1), 207–219.

Rao, R. P. (2004). Bayesian computation in recurrent neural circuits. *Neural Computation, 16,* 1–38.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 127–140.

Thura, D., Beauregard-Racine, J., Fradet, C. W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology, 108*(11), 2912–2930.

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review, 118*(4), 583–613.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550–592.

van Ravenzwaaij, D., Mulder, M. J., Tuerlinckx, F., & Wagenmakers, E. J. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. *Frontiers in Psychology*, 3.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Wagenmakers, E. J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641–671.

Wagenmakers, E. J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58,* 140–159.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics, 19*(3), 326–339.